|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Species** | **KU\_C** | **KU\_N** | **KU 70/80 beta barrel** | **VWF\_A** | **KU\_pk\_bind** |
| *Strongylocentrotus purpuratus |XP\_030843748.1* | 468-562 | 8-171 | 245-453 |  | 588-703 |
| *Nematostella vecentsis |XP\_001628737.1* | 277-372 |  | 64-241 |  |  |
| *Hydra vulgaris |XP\_012555181.1* | 460-533 | 8-194 | 241-440 |  | 575-689 |
| *Monodelphus domestica| XP\_016280585.1* | 363-456 | 1-129 | 139-338 |  | 480-592 |
| *Homo sapiens |P13010* | 477-569 |  | 253-452 | 7-249 | 595-705 |

**a.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Species** | **VWF\_A** | **Ku\_N** | **KU 70/80 beta barrel** | **KU\_pk\_bind** |
| *Ustilago maydis 521 |XP\_011391495.1* | 8-193 |  | 269-499 | 711-821 |
| *Lacaria bicoloris | XP\_001873673.1* |  |  | 74-290 | 508-626 |
| *Puccinia graminis f. sp. Tritici |KAA1113434.1* |  | 8-151 | 367-515 |  |
| *Mortierlla alpine | BAU24610.1* |  | 6-166 | 242-453 | 655-808 |
| *Schizosaccharmyces pombe |NP\_596791.1* |  |  | 240-453 | 589-687 |
| *Neurospora crassa OR74A |XP\_956387.3* | 53-214 |  | 291-496 | 651-772 |
| *Aspergillus oryzae RIB40 |Q2MHH2.1* |  | 6-189 | 241-452 | 600-717 |
| *Talaromyces marneffei PM1 |KFX43477.1|2|* |  | 6-107 | 211-421 | 563-680 |
| *Talaromyces marneffei ATCC 18224 |XP\_002151653.1|1|* |  | 6-160 | 241-449 | 593-713 |
| *Sphaeroforma arctica JP610 |XP\_014153316.1* |  |  2-95 | 165-379 | 527-645 |
| *Saccharomyces cerevisiae |Q04437* |  | 5-231 | 242-455 |  |

**b.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Species** | **Ku\_N** | **KU70/KU80 beta barrel** | **KU\_pk\_bind** |
| *Paramecium tetraurelia strain d4-2|XP\_001426860.1* | 6-133 | 250-440 | 600-711 |

**c.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Species** | **VWF\_A** | **Ku\_N** | **Ku70/Ku80 beta barrel** |
| *Acanthamoeba castellanii str. Neff |XP\_004345889.1* | 8-232 | 268-467 | 268-467 |
| *Dictyostelium discoideum AX4| XP\_637846.1* | 10-169 | 646-748 | 240-471 |

**d.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Species** | **Ku\_C** | **Ku\_N** | **KU70/KU80 beta barrel** | **KU\_pk\_bind** |
| *Arbidopsis thalina AAF79532.1 F21D18.26* | 489-557 | 45-175 | 240-444 | 585-698 |
| *Volvox carteri f. nagariensis| XP\_002954094.1* |  |  | 498-653 | 900-997 |

**e.**

Table supp. 1: Domain specifications of the predicted Ku80 protein sequences of (a) Animalia (b) Fungi (c) Ciliophora (d) Amoebozoa (e) Archaeplastidae. The sequences were obtained using NCBI BLASTP against *Homo sapiens* Ku80 query sequence. Domain specifications was obtained from InterPro.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Species** | **KU\_N** | **KU\_C** | **KU 70/80 beta barrel** | **VWF\_A** | **SAP-dom** |
| *Strongylocentrotus purpurat|XP\_030845410.1* | 38-253 | 471-505 | 262-458 | 38-244 |  |
| *Branchiostoma floridae| XP002597564.1* | 44-263 | 480-559 | 272-468 |  | 575-610 |
| *Nematostella vectensis| XP\_001641354.1* | 39-253 | 470-554 | 261-457 |  | 571-604 |
| *Hydra Vulgaris| XP\_002160929.2* | 35-250 | 467-548 | 260-455 | 35-241 |  |
| *Monodelphis domestica| XP\_001378778.1* | 37-255 | 472-557 | 264-460 | 37-246 |  |
| *Homo sapiens KU70|P12956* | 37-255 | 473-557 | 265-454 |  | 573-607 |
| *Monosiga brevicolis|* *EDQ89450.1* | 774-891 | 1104-1180 | 901-1082 |  | 1200-1234 |

**a.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Species** | **Ku\_N** | **Ku\_C** | **Ku70/80 Beta barrel** | **SAP\_dom** |
| *Malassezia globosa|XP\_001731438.1* | 43-164 |  | 246-446 |  |
| *Ustilago maydis| XP\_011388207.1* | 42-229 | 574-641 | 351-559 | 673-707 |
| *Laccaria bicolor| XP\_001878299.1* | 31-270 | 522-601 | 284-499 | 616-650 |
| *Puccinia graminis| KAA1090412.1|1|* | 63-265 | 517-587 | 290-492 |  |
| *Puccinia graminis| XP\_003335612.2|2|* |  | 143-213 | 3-118 |  |
| *Schizosaccharomyces pombe| NP\_001342782.1* | 19-236 | 457-543 | 244-436 | 570-604 |
| *Aspergillus oryzae| Q2MHH4.1* | 31-239 | 500-591 | 287-479 | 618-652 |
| *Talaromyces marneffei| XP\_002145509.1* |  |  | 1-651 |  |
| *Saccharomyces cerevisiae |A6ZN00* | 31-267 | 277-482 | 493-599 |  |

**b.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Species** | **Ku\_N** | **Ku\_C** | **Ku70/80 beta barrel** | **SAP\_Dom** | **PGBD** | **VWF\_A** |
| *Oxytricha trifallax| EJY72241.1* | 43-267 | 499-564 | 277-468 |  |  | 43-258 |
| *Stylonychia lemnae| CDW84185.1* | 43-267 | 499-569 | 277-468 |  |  |  |
| *Tetrahymena thermophila| XP\_001470738.1|1|* | 42-256 | 477-564 | 265-454 | 674-708 |  |  |
| *Tetrahymena thermophila| XP\_001025145.2|2|* | 35-255 |  | 265-465 |  |  |  |
| *Tetrahymena thermophila| DAA80465.1|3|* |  |  | 233-442 |  | 905-1287 |  |
| *Paramecium tetraurelia| XP\_001440829.1* | 40-253 | 471-522 | 262-450 |  |  |  |

**c.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Species** | **Ku\_N** | **Ku\_C** | **Ku70/80 beta barrel** | **APLF\_PBZ** |
| *Dictyostelium discoideum| XP\_637925.1* | 72-302 | 597-660 | 325-585 | 886-909 |

**d.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Species** | **Ku\_N** | **Ku\_C** | **Ku70/80 beta barrel** | **PGDB** |
| *Arabidopsis thaliana| NP\_564012.1* | 30-266 | 487-572 | 275-473 | 585-619 |
| *Volvox carteri|XP\_002947132.1* |  | 215-294 | 40-195 | 325-359 |
| *Ostreococcus lucimarinus|XP\_001419207.1* |  | 173-212 | 3-154 | 298-332 |

**e.**

Table 2: Domain specifications of predicted Ku70 protein sequences of (a) Animalia (b) Fungi (c) Ciliophora (d) Amoebozoa (e) Archaeplastidae. The sequences were obtained using NCBI BLASTP against *Homo sapiens* Ku70 query sequence. Domain specifications was obtained from InterPro.

|  |  |  |
| --- | --- | --- |
| Animalia |  |  |
| **S.No.** | **Organism** | **Genome Sequenced** |
| 1. | *Hydra magnipapillata* | None |
| 2. | *Buddenbrockia plumatellae* | None |
| 3. | *Branchiostoma floridae* | Partial |
| 4. | *Strongylocentrotus purpurat* | Partial |
| 5. | *Nematostella vectensis* | Partial |
| 6. | *Homo sapiens* | Partial |
| 7. | *Monodelphis domestica* | Partial |
| Fungi |  |  |
| **S.No.** | **Organism** | **Genome Sequenced** |
| 1. | *Malassezia globosa* | Partial |
| 2. | *Candida albicans* | Partial |
| 3. | *Moniliopthora perniciosa* | None |
| 4. | *Neurospora crassa* | Partial |
| 6. | *Mortierella alpina* | Partial |
| 7. | *Ustilago maydis* | Partial |
| 8. | *Laccaria bicolor* | Partial |
| 9. | *Puccinia graminis* | Partial |
| 10. | *Schizosaccharomyces pombe* | Partial |
| 11. | *Penicillium marneffei* | Partial |
| 12. | *Saccharomyces cerevisiae* | Complete |
| 13. | *Sphaeroforma arctica* | Partial |
| 14. | *Aspergillus oryzae* | Partial |
| Amoebozoa |  |  |
| **S.No.** | **Organism** | **Genome Sequenced** |
| 1. | *Entamoeba invadens* | Partial |
| 2. | *Entamoeba histolytica* | Partial |
| 3. | *Physarum polycephalum* | Partial |
| 4. | *Hyperamoeba dachnya* | None |
| 5. | *Mastigamoeba balamuthi* | Partial |
| 6. | *Acanthamoeba castellanii* | Partial |
| 7. | *Dictyostelium discoideum* | Partial |
| Ciliophora |  |  |
| **S.No.** | **Organism** | **Genome Sequenced** |
| 1. | *Oxytricha trifallax* | Partial |
| 2. | *Euplotes crassus* | Partial |
| 3. | *Euplotes octocarinatus* | None |
| 4. | *Anophryoides haemophila* | None |
| 5. | *Stylonychia lemnae* | Partial |
| 6. | *Tetrahymena thermophila* | Partial |
| 7. | *Paramecium caudatum* | Partial |
| 8. | *Paramecium tetraurelia* | Partial |
| Archaeplastidae |  |  |
| **S.No.** | **Organism** | **Genome Sequenced** |
| 1. | *Griffithsia japonica* | None |
| 2. | *Cyanidioschyzon merolae* | Complete |
| 3. | *Chlamydomonas reinhardtii* | Partial |
| 4. | *Scenedesmus obliquus* | Partial |
| 5. | *Acetabularia acetabulum* | None |
| 6. | *Micromonas* | Partial |
| 7. | *Glaucocystis nostochinearum* | None |
| 8. | *Cyanophora paradoxa* | Partial |
| 9. | *Ostreococcus lucimarinus* | Complete |
| 10. | *Arabidopsis thaliana* | Partial |
| 11. | *Volvox carteri* | Partial |
| Eozoa |  |  |
| **S.No.** | **Organism** | **Genome Sequenced** |
| 1. | *Trypanosoma cruzi* | Partial |
| 2. | *Trypanosoma brucei* | Partial |
| 3. | *Leishmania infantum* | Partial |
| 4. | *Leishmania major* | Complete |
| 5. | *Diplonema papillatum* | Partial |
| 6. | *Giardia lamblia* | Partial |
| 7. | *Spironucleus vortens* | None |
| 8. | *Streblomastix strix* | Partial |
| 9. | *Euglena gracilis* | Partial |
| 10. | *Sawyeria marylandensis* | None |
| 11. | *Trichomonas vaginalis* | Partial |
| 12. | *Reclinomonas americana* | None |

Table 3: Genome sequencing status of all 59 selected eukaryotes. Absence of Ku sequence during BLASTP in some eukaryotes might be due to its incompletely sequenced genome.



Bacteriodetes

Acidobacteria

Archaeplastidae

Delta-epsilon proteobacteria

Verrucomicrobia

Fungi

Ciliophora

Amoeboza

Fungi

Animalia

Archaeplastidae

Fungi

Archaeplastidae

Bacteriodetes

Ciliophora

Archaeplastidae

Choanoflagellates

Fungi

Amoeboza

Fungi

Animalia

Euyarchaeota

Fungi

Actinoplanes

Alphaproteobacteria

Thermodesulfobacter

Euyarchaeota

Betaproteobacteria

Delta-epsilon proteobacteria

Firmicutes

Euyarchaeota

Acidobacteria

Verrucomicrobia

Betaproteobacteria

Figure 1: Phylogenetic tree inferred using Neighbor-Joining method(N. & M., 1987).  The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches(Felsenstein, 1985). This analysis involved 84 amino acid sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). Evolutionary analyses were conducted in MEGA X(Kumar et al., 2018). Organisms are labelled according to their respective phyla/kingdoms on the left of the tree.



Acidobacteria

Alphaproteobacteria

Thermodesulfobacter

Euyarchaeota

Delta-epsilon proteobacteria

Betaproteobacteria

Verrucomicrobia

Betaproteobacteria

Verrucomicrobia

Bacteriodetes

Delta-epsilon proteobacteria

Actinobacteria

Acidobacteria

Firmicutes

Euyarchaeota

Actinobacteria



Animalia

Fungi

Fungi

Amoeboza

Archaeplastidae

Firmicutes

Archaeplastidae

Fungi

Ciliophora

Fungi

Ciliophora

Actinobacteria

Choanoflagellates

Archaeplastidae

Amoeboza

Animalia

Archaeplastidae

Ciliophora

Fungi

Euyarchaeota

Fungi

Figure 2. *Streptomyces* species; S. ambofaciens, S. *griseus*, S. *cattleya*, S. *coelicolor*, S. *albulus* and S. *avermitilis* (Hoff et al., 2016) and *Mycobacterium tuberculosis* Ku sequences were added and phylogenetic tree was inferred by using the Maximum Likelihood method and Whelan And Goldman model(Whelan & Goldman, 2001) in MEGA-X. The bootstrap consensus tree inferred from 500 replicates (Felsenstein, 1985) is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. Substitution model used was WAG model. Gamma Distributed with Invariant sites (G+I) rates was used with 5 number of discrete gamma categories. Partial deletion was used for Gaps/Missing data treatment with 95% site coverage cutoff. Nearest-Neighbor-Interchange (NNI) ML Heuristic method was used and Make initial tree automatically (Default – NJ/BioNJ) was chosen for Initial Tree for ML. Branch Swap filter was none and Number of threads selected was 3 (Hall, 2013). Corresponding phyla/kingdoms are labelled at left of the tree.

References

Felsenstein, J. (1985). Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution*. https://doi.org/10.2307/2408678

Hall, B. G. (2013). Building phylogenetic trees from molecular data with MEGA. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/mst012

Hoff, G., Bertrand, C., Zhang, L., Piotrowski, E., Chipot, L., Bontemps, C., Confalonieri, F., McGovern, S., Lecointe, F., Thibessard, A., & Leblond, P. (2016). Multiple and variable NHEJ-like genes are involved in resistance to DNA damage in Streptomyces ambofaciens. *Frontiers in Microbiology*. https://doi.org/10.3389/fmicb.2016.01901

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msy096

N., S., & M., N. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*. https://doi.org/10.1093/oxfordjournals.molbev.a040454

Whelan, S., & Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular Biology and Evolution*. https://doi.org/10.1093/oxfordjournals.molbev.a003851