CIPRO 2.5: *Ciona intestinalis* Protein integrated database with large-scale omics data, bioinformatic analyses and curated annotation, with ability for user rating and comments

Toshinori Endo[1], Keisuke Ueno[1], Kouki Yonezawa[1], Katsuhiko Mineta[1], Kohji Hotta[2], Yutaka Satou[3], Lixy Yamada[4], Michio Ogasawara[5], Hiroki Takahashi[6], Ayako Nakajima[7], Mia Nakachi[7], Mamoru Nomura[7], Junko Yaguchi[7], Alu Konno[7], Yasunori Sasakura[7], Akiyasu C. Yoshizawa[8], Hisaaki Taniguchi[8], Chisato Yamasaki[9], Miho Sera[9], Tadashi Imanishi[9] and Kazuo Inaba[7]

[1]Graduate School of Information Science and Technology, Hokkaido University, Sapporo 060-0814, Japan.
[2] Department of Biosciences and Informatics, Keio University, Yokohama 223-8522, Japan.
[3] Department of Zoology, Graduate School of Science, Kyoto University, Sakyo, Kyoto 606 8502, Japan.
[4] Sugashima Marine Biological Laboratory, Nagoya University, 429-63 Sugashima, Toba 517-0004, Japan.
[5] Department of Biology, Faculty of Science, Chiba University, Inage-Ku, Chiba 263-8522, Japan.
[6] National Institute for Basic Biology, Nishigonaka 38, Myodaiji, Okazaki 444-8585, Japan.
[7] Shimoda Marine Research Center, University. of Tsukuba, 5-10-1 Shimoda, Shizuoka, 415-0025, Japan.
[8] Institute for Enzyme Research, the University of Tokushima, 3-18-15, Kuramoto-cho, Tokushima 770-8503, Japan.
[9] Biomedicinal Information Research Center, Advanced Industrial Science and Technology, Waterfront Bio-IT Research Building, 2-42 Aomi, Koto-ku, Tokyo 135-0064, Japan.

CIPRO database is an integrated protein database for a tunicate species Ciona intestinalis that belongs to the Urochordata. Although the CIPRO data deals with proteomic and transcriptomic data of a single species, the animal is considered unique in the evolutional tree, representing a possible origin of the vertebrates and is a good model for understanding chordate evolution, including human. Furthermore, C. intestinalis has been one of the favorites of developmental biologists; there exist huge amount of accumulated knowledge on its development, morphology, in addition to the recent genome sequence and gene expression data. CIPRO database is aimed at not only collecting those published data, but also presenting unique information including the unpublished transcriptomic and proteomic data and human curated annotation, for the use by researchers in broad research fields of biology and bioinformatics.

The current database contains 89,683 unique sequences covering all the proteins from all the gene models on this species; the number was reduced to 70,493 by similarity clustering. these sequences, more than 5,000 proteins are manually annotated based on the large-scale transcriptomic, proteomic and bioinformatic data. annotations can be subjected to be qualification by rating, curation, and comments by named and anonymous users through the web site of CIPRO database. views are shown in figures 1 and 2.

Unique features of CIPRO database include:

1. Original experimental data

Unpublished experimental data, including 2D-PAGE with the identified protein spots by protein mass fingerprint (PMF) MS analysis, expressions or localizations of protein and RNA across developmental stages and tissues, altogether summarized in a single chart for the comparison among status and methods. RNA expressions are observed by microarray and EST. Each protein is linked to an independent Ascidian Proteome Database summarizing large-scale MS-based proteomic analyses.

2. Whole *Ciona intestinalis* proteome database

Proteins across gene models are presented: all protein models derived from published gene models are incorporated, including Kyoto model (KG), KH (successor of KG model), PROCITS, JGI's versions 1 and 2, and Ensembl (version 58.2) are incorporated. Identical sequences across gene models are shown.

3. Original comprehensive user-friendly interfaces

Bioinformatic analyses and prediction results are summarized in pictures for grasp at a glance: homology search, cytolocalization, secondary structure prediction combined with modification sites, such include phosphorylation and three-dimensional structures.

4. Comparative analysis data for disease association

Comparison with human genome: map location of human homologues is graphically shown with associated disease information. Comparative data for other model organisms are also included.

5. Community-wide curation capability opened to users

To facilitate progressive improvement of annotation by visited users, users can place additional annotation for the protein name and/or comments, which will be subjected to rating by the followed data viewers. To aid curation by wide community, information for literature and essence of matched motif patterns and other related protein information are shown with the links.

6. Useful search facilities

Various search methods are provided including blast homology, free text, partial sequence, protein mass fragment, and cross item searches.

7. Experimental assistance information

For experimental design, possibly interacting antibody products are suggested. In addition, information of the chemicals affecting the tunicate morphology or development are also provided.

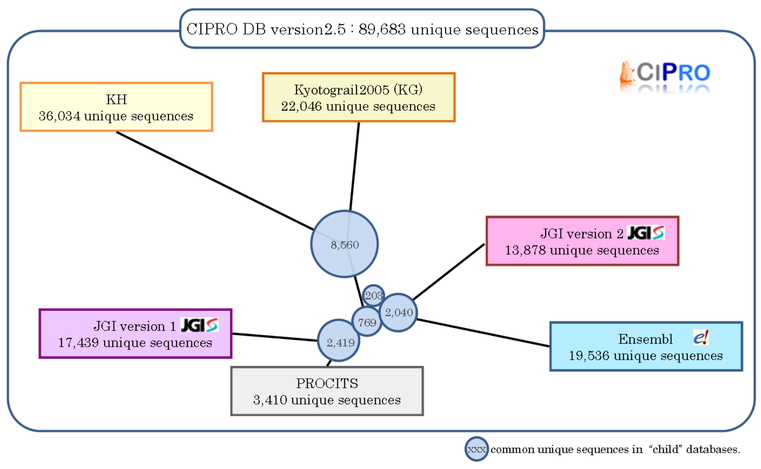
Site URL

CIPRO is available at http://cipro.ibio.jp/2.5 .

Usage statistics

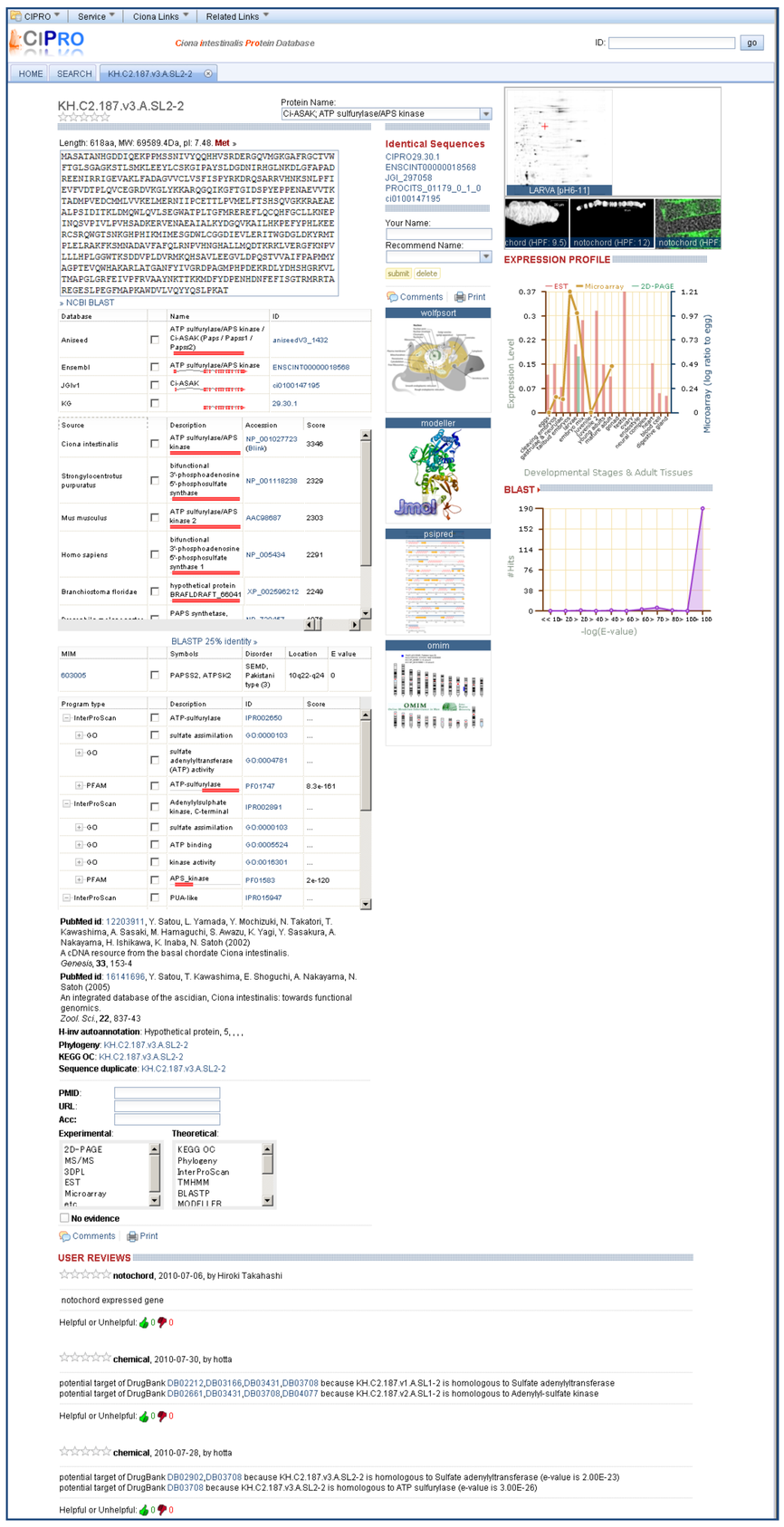
CIPRO has been visited by 68,582 users for 2,582,236 page views in the last 12 month as of June 30, 2010.

CIPRO Sequence Data Composition



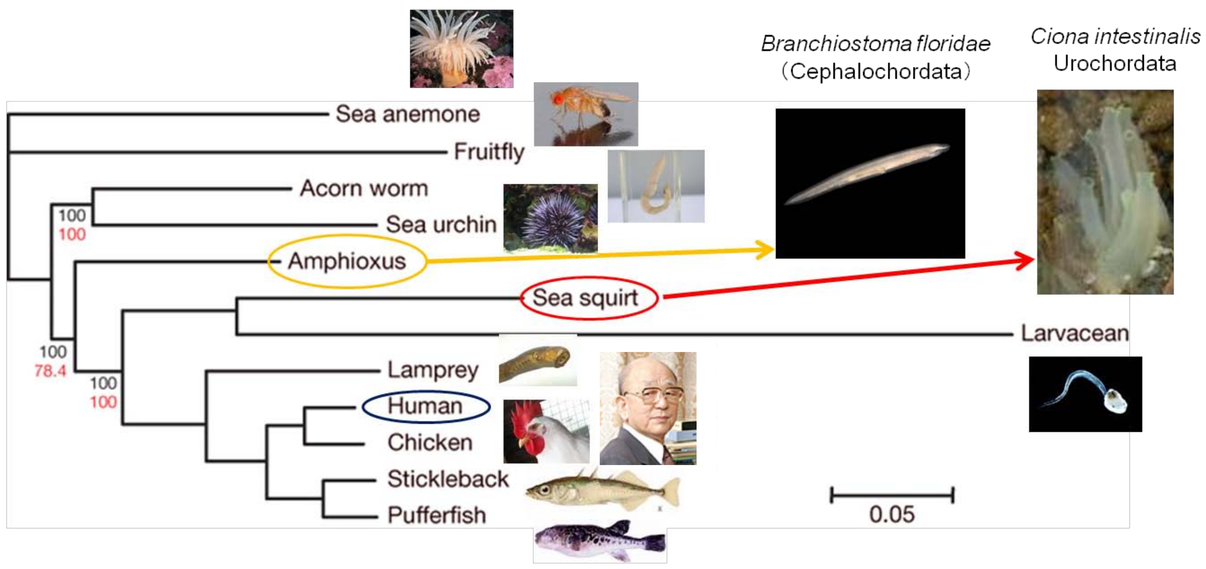
Five proteomes and one dataset are contained in CIPRO. The branch length represents the distance of the database . As can be seen, they share small proportion of sequences each others.

Protein Information

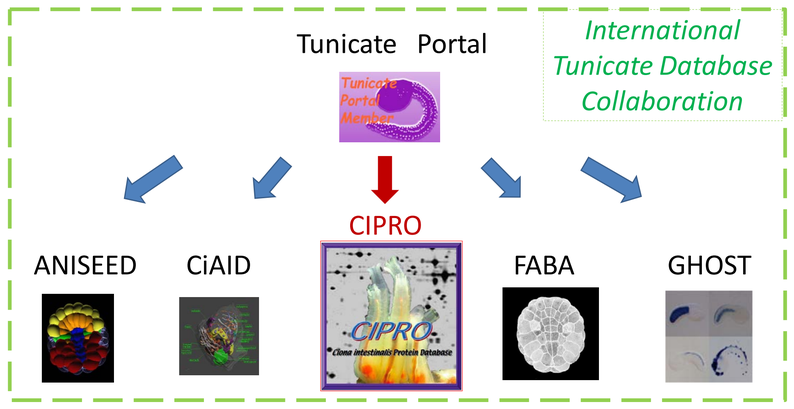


Text and graphic data are integrated to be self-explanatory for the protein. A visitor can add a comment on the protein.

Urochodates are the closest relatives to vertebrates



Modified from Putnam et al. Nature 453, 1064-1071 (19 June 2008)



Nature Precedings : doi:10.1038/npre.2010.5435.1 : Posted 24 Dec 2010