



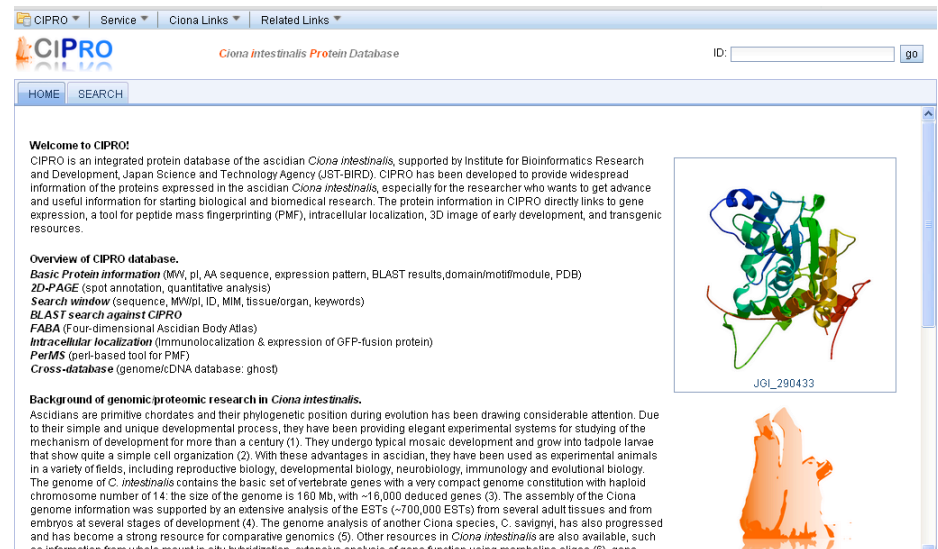
Incorporating Community Annotation Interfaces into the CIPRO2.5 Database with Comprehensible Sketches to Support Quick Annotations of Proteome Data.

Keisuke Ueno, Kouki Yonezawa Katsuhiko Mineta,
Toshinori Endo

Graduate School of Information Science and
Technology, Hokkaido University, North 14 West
9, Sapporo, Hokkaido 060-0814, Japan

The system for wet laboratory workers is the goal of *Ciona intestinalis* protein database (CIPRO)

- To improve usability for experimental ascidian researchers
- To give feedback to the ascidian researchers
- To facilitate interactions with other researchers concerned with developmental biology, evolutionary biology and human diseases



The screenshot displays the CIPRO (Ciona intestinalis Protein Database) website. At the top, there are navigation tabs for 'CIPRO', 'Service', 'Ciona Links', and 'Related Links'. Below the tabs is the CIPRO logo and the text 'Ciona intestinalis Protein Database'. A search bar with an 'ID:' label and a 'go' button is located on the right. The main content area features a 'HOME' and 'SEARCH' button. The 'Welcome to CIPRO!' section provides an overview of the database, supported by the Institute for Bioinformatics Research and Development, Japan Science and Technology Agency (JST-BIRD). It lists various services available, such as Basic Protein Information (MW, pI, AA sequence, expression pattern, BLAST results, domain/motif/module, PDB), 2D-PAGE, Search window, BLAST search against CIPRO, FABA, Intracellular localization, PerMS, and Cross-database. A 3D protein structure visualization is shown on the right, labeled 'JGI_290433'. Below the 3D structure is a 2D gel electrophoresis image. The 'Background of genomic/proteomic research in Ciona intestinalis' section discusses the organism's phylogenetic position, developmental process, and the genome's characteristics, including its compact size and the number of genes.

Biological database requires several basic functionalities

- To evaluate reliability, to keep quality, and to provide timely information
- To reduce redundancy and to synchronize cross-references
- To lead toward desired information by using retrieval systems and providing reticulated links rather than ordinary hierarchical links

Community websites & community annotation become the next standard for data resources?

- Social network sites have grown in importance globally
- Wikipedia is also a successful example. However, the reliability of descriptions is controversial
- The community annotation among experts is ideal



WIKIPEDIA
The Free Encyclopedia

Incorporating three new functions into CIPRO2.5 database

- Formatted web forms & a free comment editor
- Comprehensible sketches
- Retrieval systems including combined fields

Two types of annotation interfaces make up for each other's deficiencies

- Web form

- Comment editor

•Checkbox

Protein Name:

Length: 161aa, mol. wt. 16630.00Da, pI 5.14, mol. wt.

MMRFMLLFSROGK YRLQKWFNAIGDVERKKITRELVSLVLSRRPKMCFDWDKDYK
VYYKRYASLFFCAIEESDNELLAEVIRHFVEILDKYFGSYCELDIIFNFEKAYF
ILDEFLLAGEIQESSRNKYLRAVENS DALQENDEAEATRSILEEFGLS

» NCBI BLAST

Database	Name	ID
Aniseed	<input type="checkbox"/> (AP1S2 / AP-1sigma / Ap1s1 / SOX12 / Coq9)	aniseedV3_1330
Ensembl	<input type="checkbox"/>	ENSCINT00000025653

Source	Description	Accession	Score
	PREDICTED: similar to AP-1 complex subunit sigma-2 (Adapter-related protein complex 1 sigma-1B subunit) (Adaptor protein		
		2122331	832
	sigma-1B subunit (Clathrin assembly pr... isoform 2		
Mus musculus	<input type="checkbox"/> unnamed protein product	BAC31652	649

•Dropdown list

Identical Sequences

- CIPRO38.38.1
- ENSCINT0000011259
- ENSCINT00000025653
- JGI_204207
- JGI_206968

Your Name:

Recommend Name:

Comments

wolfpsort

modeller

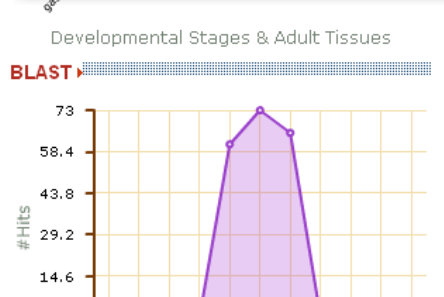
CIPRO *Ciona intestinalis* Protein Database

HOME SEARCH JGI_290433 New Comment #1

CIPRO ID: Comment Title:

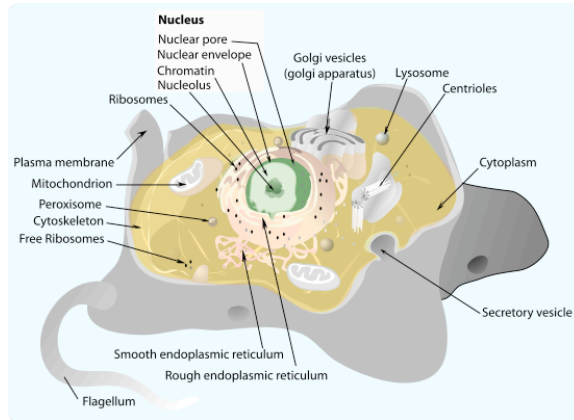
Rating: ☆☆☆☆ Your Name:

•HTML text with links and images

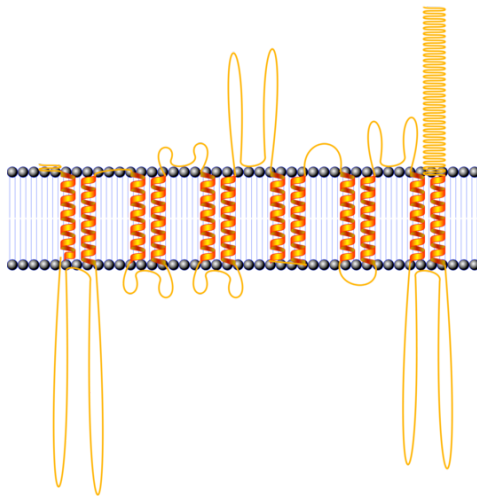


Comprehensible sketches were designed to be compact for quickly understandable overview

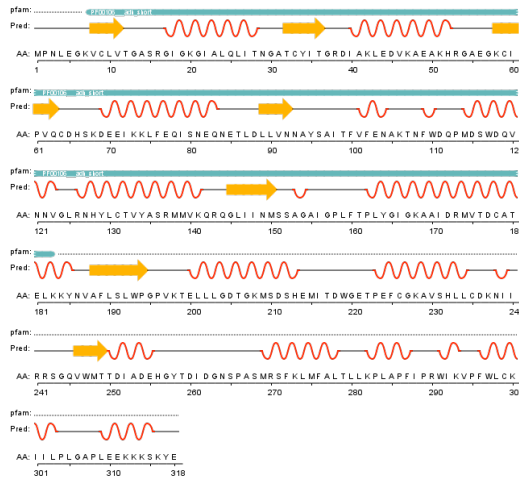
Cytolocalization



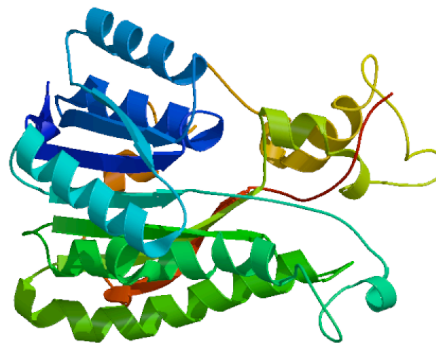
TM prediction



Sequence & motifs

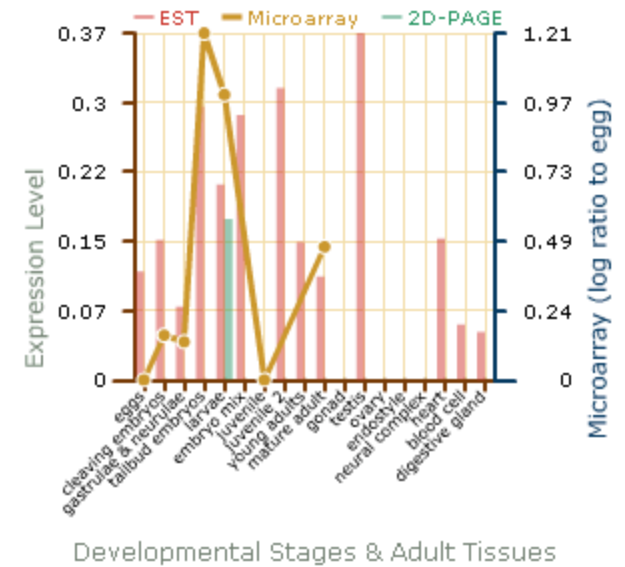


Homology model



Biocuration 2010

Expression data



Developmental Stages & Adult Tissues

Retrieval system with combined fields is useful for searching a familiar protein and its neighbors

The image shows a screenshot of the CIPRO database interface. The main window is titled "CIPRO Ciona intestinalis Protein Database". It features a navigation bar with "HOME", "SEARCH", and "JGI_290433". Below this, there are two main search sections: "General Search" and "Peptide Search".

The "General Search" section includes fields for "Full text search", "Annotator", "#Annotation" (with min and max values), "H-inv autoannotation", "Length [AA]", "MW [Da]", "pI", "Indirectly Search by Homolog", "MIM ID" (selected) or "Accession Number", and "Tissue/Stage" (with a dropdown menu showing options like "eggs", "cleaving embryos", "gastrulae neurulae", "tailbud embryos", "larvae").

The "Peptide Search" section includes fields for "Fragment Sequence", "PMF", "Modifications" (with a dropdown menu showing "N-terminal Acetylation", "Oxidation of M", etc.), "S-alkyl reagent" (with radio buttons for "None", "c", "m", "e"), "Threshold" (set to 0.2), "Contaminants", "Keratin", and "Trypsin" (with radio buttons for "None", "Porcine(Major)", "Porcine(All)", "Bovine").

Two blue boxes highlight specific areas: one labeled "Full" points to the "Full text search" field, and another labeled "PMF" points to the "PMF" field in the "Peptide Search" section.

Case study: annotation of protein names by using web forms

CIPRO Service Ciona Links Related Links

CIPRO *Ciona intestinalis* Protein Database ID:

HOME SEARCH JGI_204207 KH.C10.42.v1.A.SL1-1

KH.C10. **•Check the evidence list**

Length: 161aa, MW: 18000.00Da, pI: 5.17, mol wt

MMRFMLLFSRQGVRLQKWFNAIGDKERKKITRELYSLVLSRRPKMCFNLDWQDYK
 VYVKRYASLFFCCAIIEESDNELLAKVYIHRFVEILDKYFGSVCELDIIFNFEKAYF
 ILDFLLAGEIQESSKNKYLRAVENSALQEEQEAETRSILEEFGLS

» NCBI BLAST

Database	Name	ID
Aniseed	<input type="checkbox"/> (AP1S2 / AP-1sigma / Ap1s1 / SOX12 / Coq9)	aniseedv3_1330
Ensembl	<input type="checkbox"/>	ENSCINT00000025653

Source	Description	Accession	Score
	PREDICTED: similar to AP-1 complex subunit sigma-2 (Adaptor-related protein complex 1 sigma-1B subunit) (Adaptor protein)		

submit delete

Comments Print

wolfsort

modeller

Mus musculus unnamed protein product BAC31652 649

EXPRESSION PROFILE

Developmental Stages & Adult Tissues

BLAST

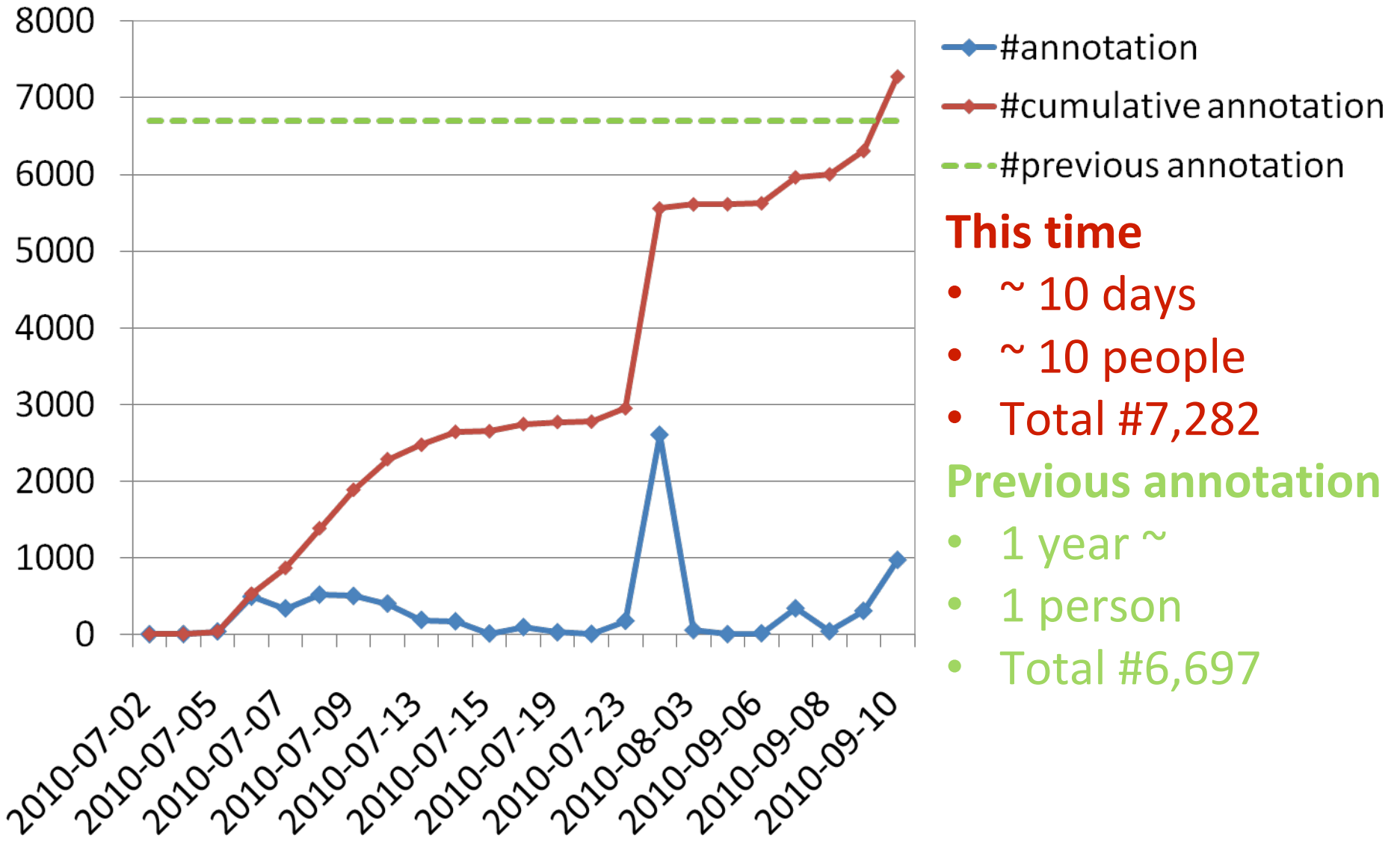
•Select a protein name

•Edit the protein name

How to decide explicit formalism against implicit criteria

- **Class I:** $\geq 50\%$ identity, $\geq 50\%$ coverage
 - HOMOLOGOUS TO
- **Class II:** $\geq 25\%$ identity
 - SIMILAR TO
- **Class III:** Found a motif or domain in databases
 - XXX domain containing proteins
- **Class IV:** Predicted proteins with evolutionary conservation
 - Conserved hypothetical proteins
- **Class V:** Predicted proteins longer than or equal to 80 aa
 - Hypothetical proteins
- **Class VI:** Predicted proteins shorter than 80 aa
 - Hypothetical short proteins

Results: Comparison of the annotation of protein names with previous annotation



This time

- ~ 10 days
- ~ 10 people
- Total #7,282

Previous annotation

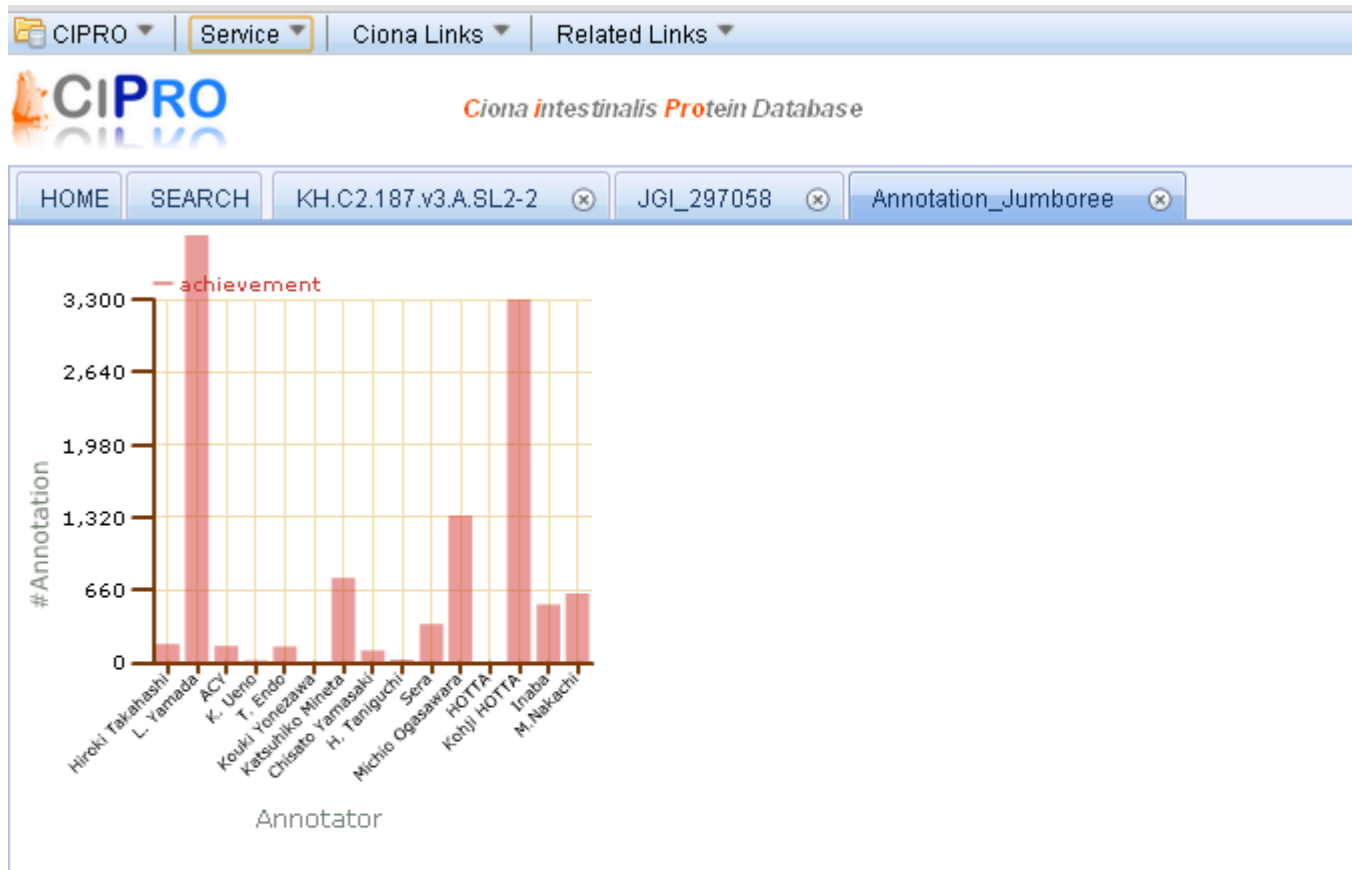
- 1 year ~
- 1 person
- Total #6,697

Flexible descriptions were essentially inescapable

- poly(A)
- poly-A
- polyadenylation

- Dehydrogenase/reductase SDR family member, short-chain dehydrogenase/reductase
- Dehydrogenase-reductase SDR family member 11

Simple chart provided an incentive



An example of useful comments by using the comment editor

The screenshot shows the CIPRO database interface for the protein *Ciona intestinalis* KH.C2.187.v3.A.SL2-2. The page includes a navigation bar with 'HOME', 'SEARCH', and tabs for 'JGI_290433', 'New Comment #1', and 'KH.C2.187.v3.A.SL2-2'. A 'No evidence' checkbox is present at the top left. Below the navigation, there are links for 'Comments' and 'Print'. The 'USER REVIEWS' section contains three entries, each with a 5-star rating, a date, and a reviewer name. The first review by 'notochord' (2010-07-06) states 'notochord expressed gene'. The second review by 'chemical' (2010-07-30) provides two potential drug targets with their DrugBank IDs and the reason for being a target (homology to sulfate adenylyltransferase and adenylyl-sulfate kinase). The third review by 'chemical' (2010-07-28) provides two potential drug targets with their DrugBank IDs and the reason for being a target (homology to sulfate adenylyltransferase and ATP sulfurylase). Each review includes a 'Helpful or Unhelpful' rating system with thumbs up and down icons and a count of 0 for both.

No evidence

[Comments](#) | [Print](#)

USER REVIEWS

☆☆☆☆☆ **notochord**, 2010-07-06, by Hiroki Takahashi

notochord expressed gene

Helpful or Unhelpful: 0 0

☆☆☆☆☆ **chemical**, 2010-07-30, by hotta

potential target of DrugBank [DB02212](#),[DB03166](#),[DB03431](#),[DB03708](#) because KH.C2.187.v1.A.SL1-2 is homologous to Sulfate adenylyltransferase
potential target of DrugBank [DB02661](#),[DB03431](#),[DB03708](#),[DB04077](#) because KH.C2.187.v2.A.SL1-2 is homologous to Adenylyl-sulfate kinase

Helpful or Unhelpful: 0 0

☆☆☆☆☆ **chemical**, 2010-07-28, by hotta

potential target of DrugBank [DB02902](#),[DB03708](#) because KH.C2.187.v3.A.SL2-2 is homologous to Sulfate adenylyltransferase (e-value is 2.00E-23)
potential target of DrugBank [DB03708](#) because KH.C2.187.v3.A.SL2-2 is homologous to ATP sulfurylase (e-value is 3.00E-26)

Helpful or Unhelpful: 0 0

YOUR RECENT HISTORY

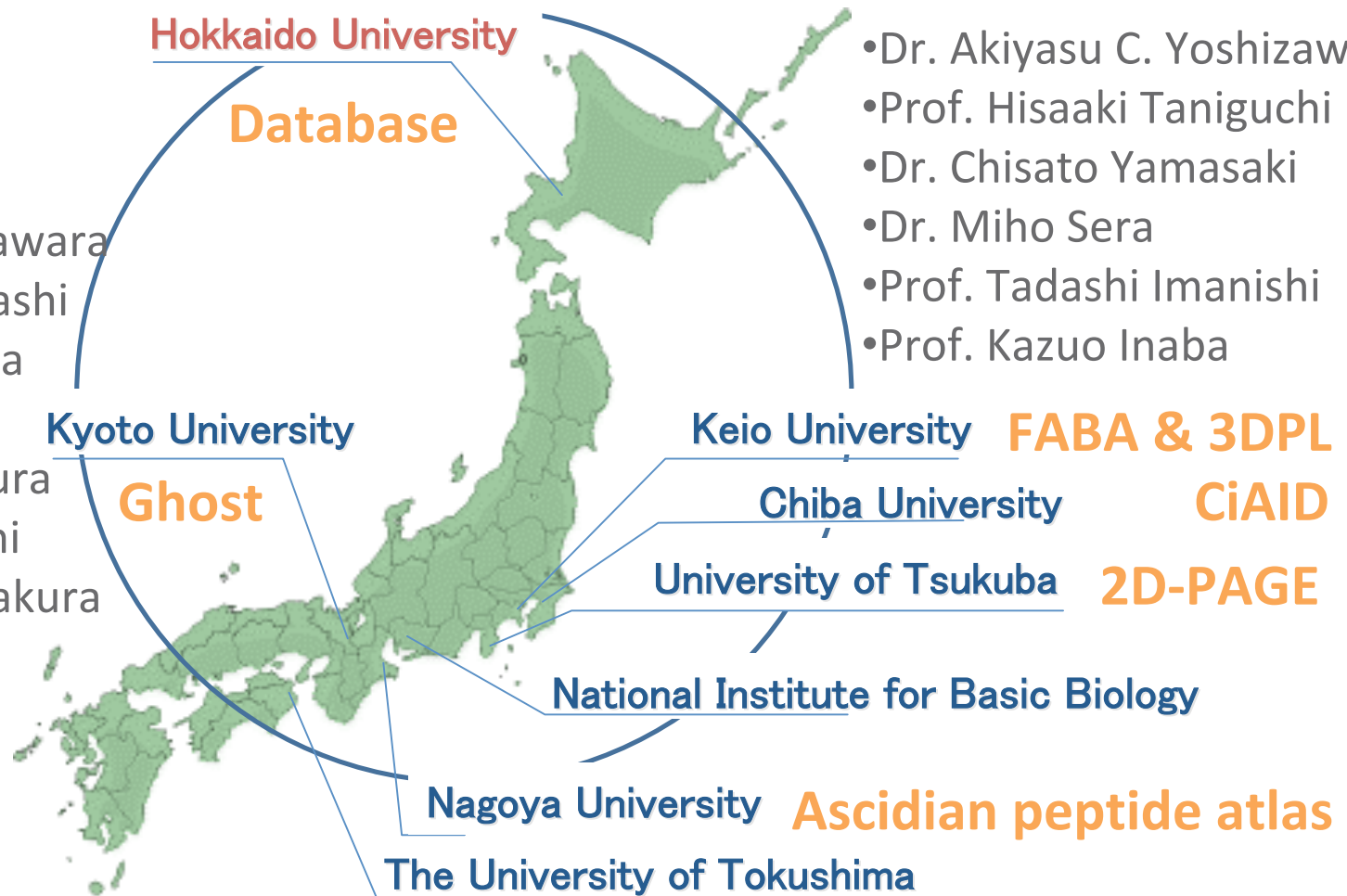
Summary

- The annotation of protein names were assisted by using formatted web forms .
- Additional information such as chemical targets were also provided by using the user comment editor.
- These annotated data are freely accessible at the CIPRO2.5 web site

Acknowledgments

Colleague

- Prof. Kohji Hotta
- Prof. Yutaka Satou
- Prof. Lixy Yamada
- Prof. Michio Ogasawara
- Prof. Hiroki Takahashi
- Dr. Ayako Nakajima
- Dr. Mia Nakachi
- Dr. Mamoru Nomura
- Prof. Junko Yaguchi
- Prof. Yasunori Sasakura



- Dr. Akiyasu C. Yoshizawa
- Prof. Hisaaki Taniguchi
- Dr. Chisato Yamasaki
- Dr. Miho Sera
- Prof. Tadashi Imanishi
- Prof. Kazuo Inaba



- This work was as supported by BIRD of Japan Science and Technology Agency (JST).