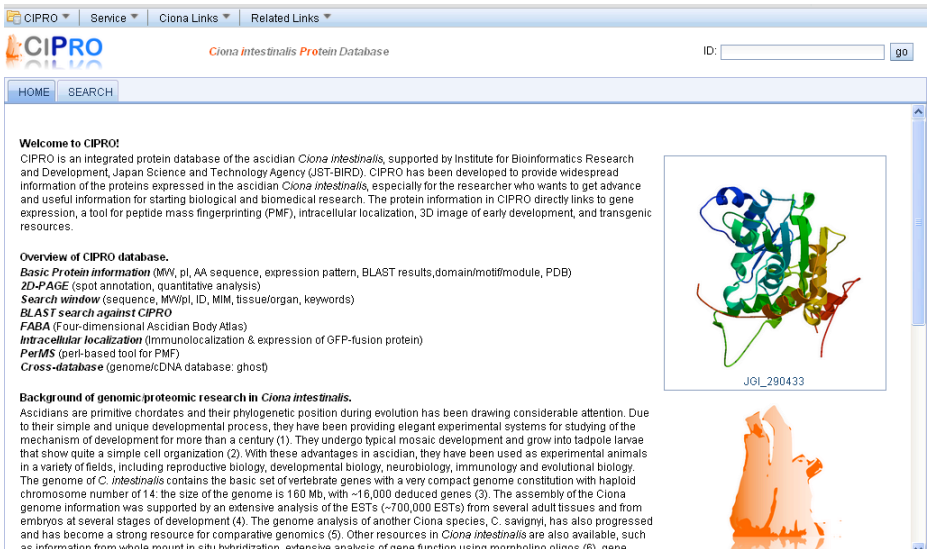
Incorporating Community Annotation Interfaces into the CIPRO2.5 Database with Comprehensible Sketches to Support Quick Annotations of Proteome Data.

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



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?

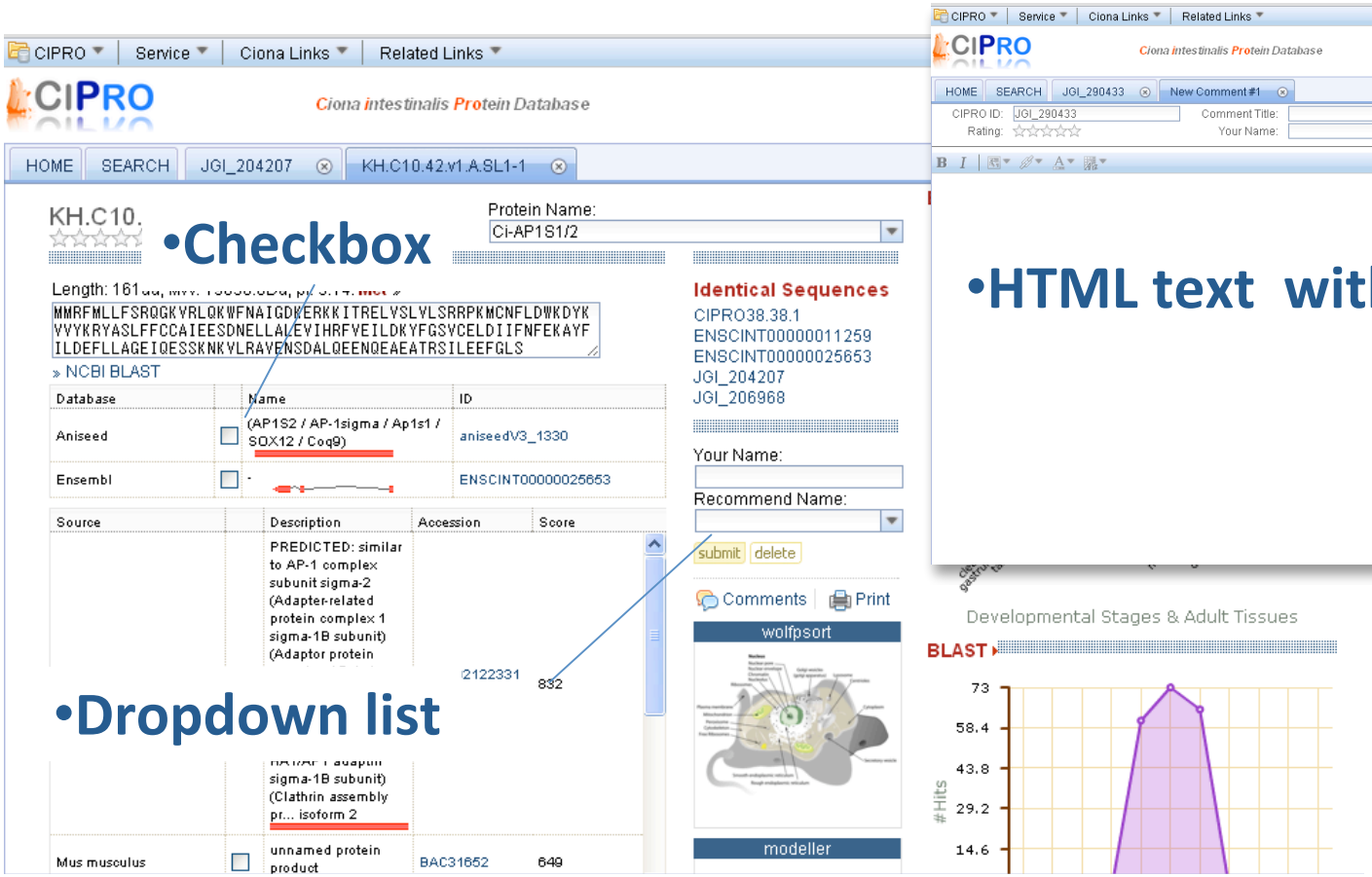
- Social network sites have grown in importance globally
- Wikipedia is also a successful example. However, the reliability of descriptions is controversial
- The community annotation among experts is ideal

Incorporating three new functions into CIPRO2.5 database

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

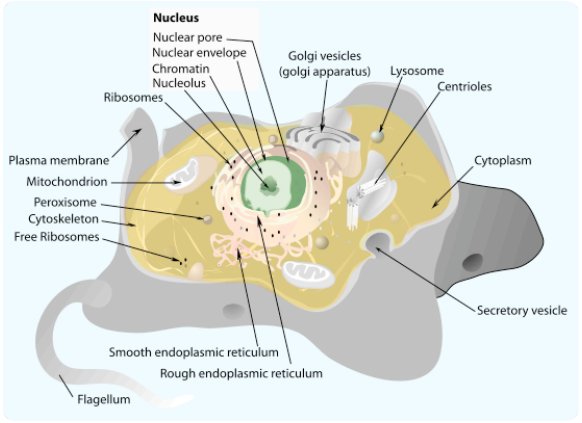
Two types of annotation interfaces make up for each other's deficiencies

- Web form
- Comment editor

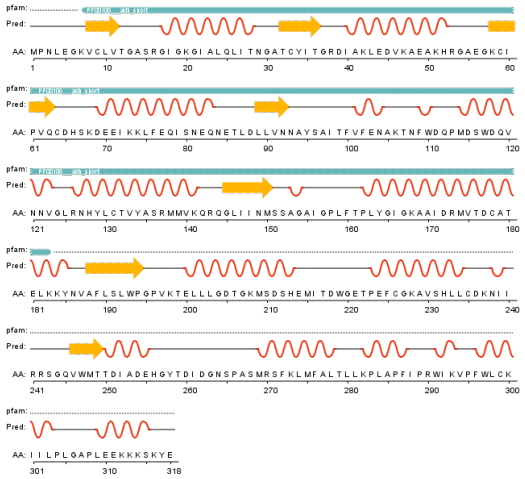


Comprehensible sketches were designed to be compact for quickly understandable overview

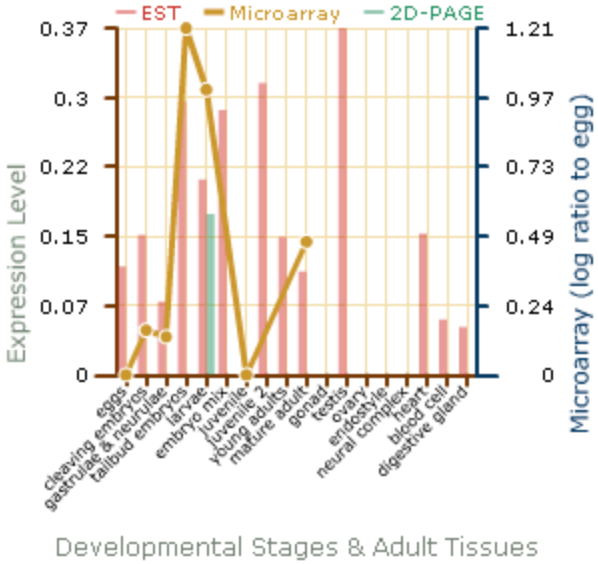
Cytolocalization



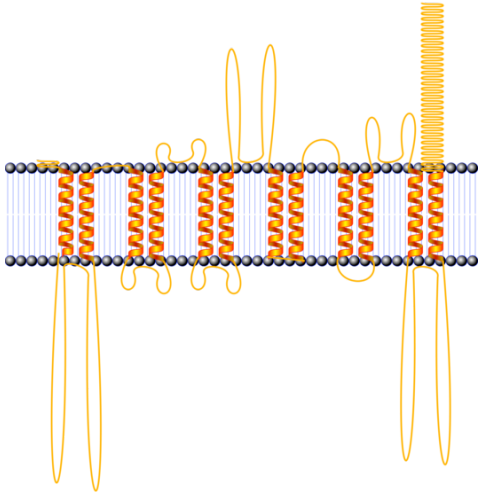
Sequence & motifs



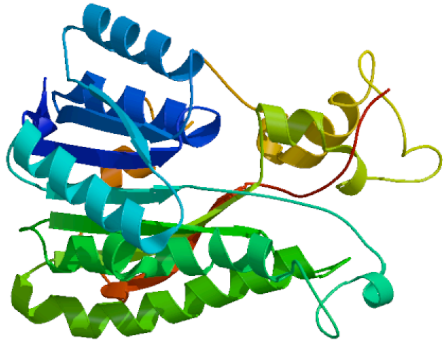
Expression data



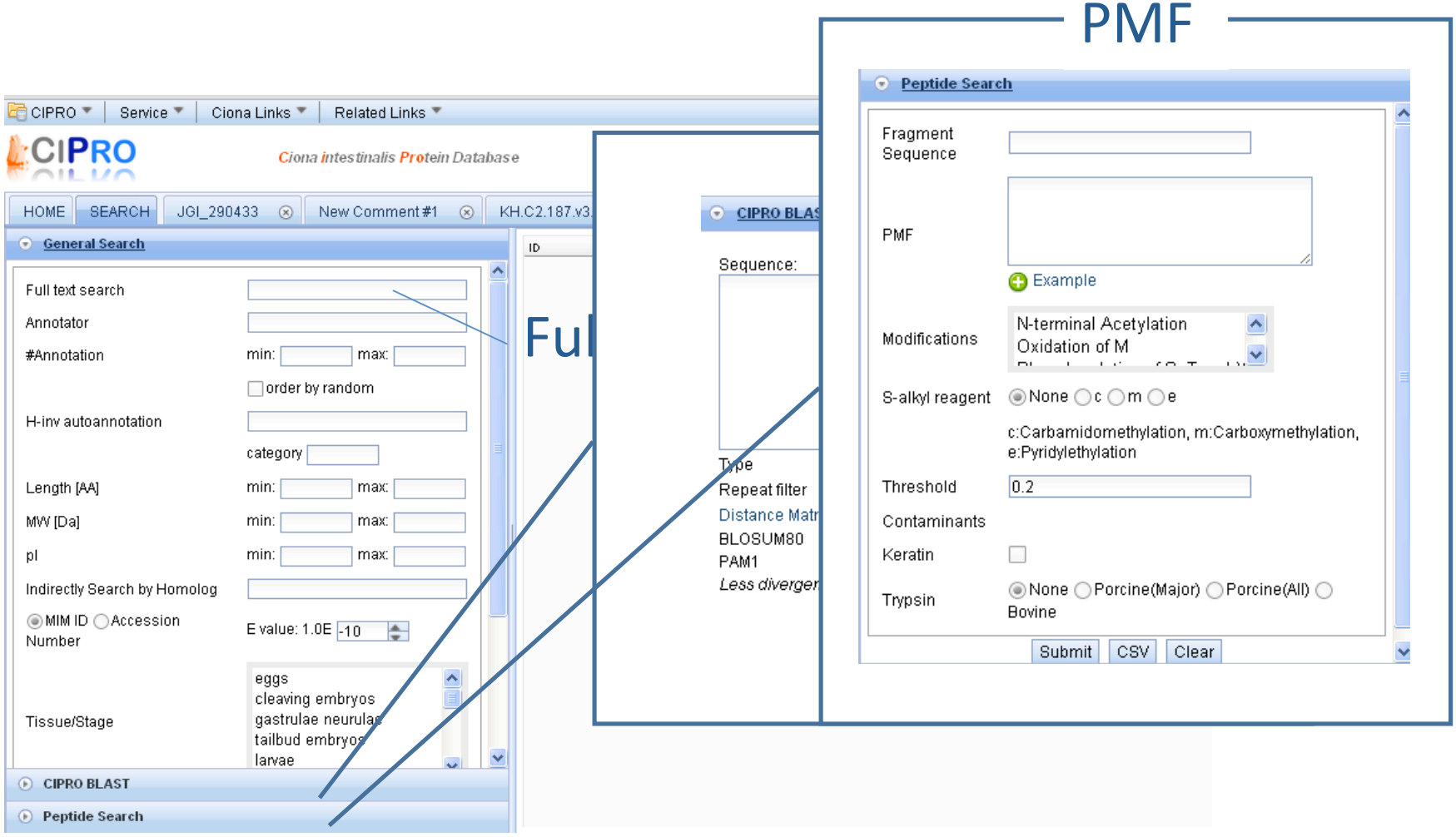
TM prediction



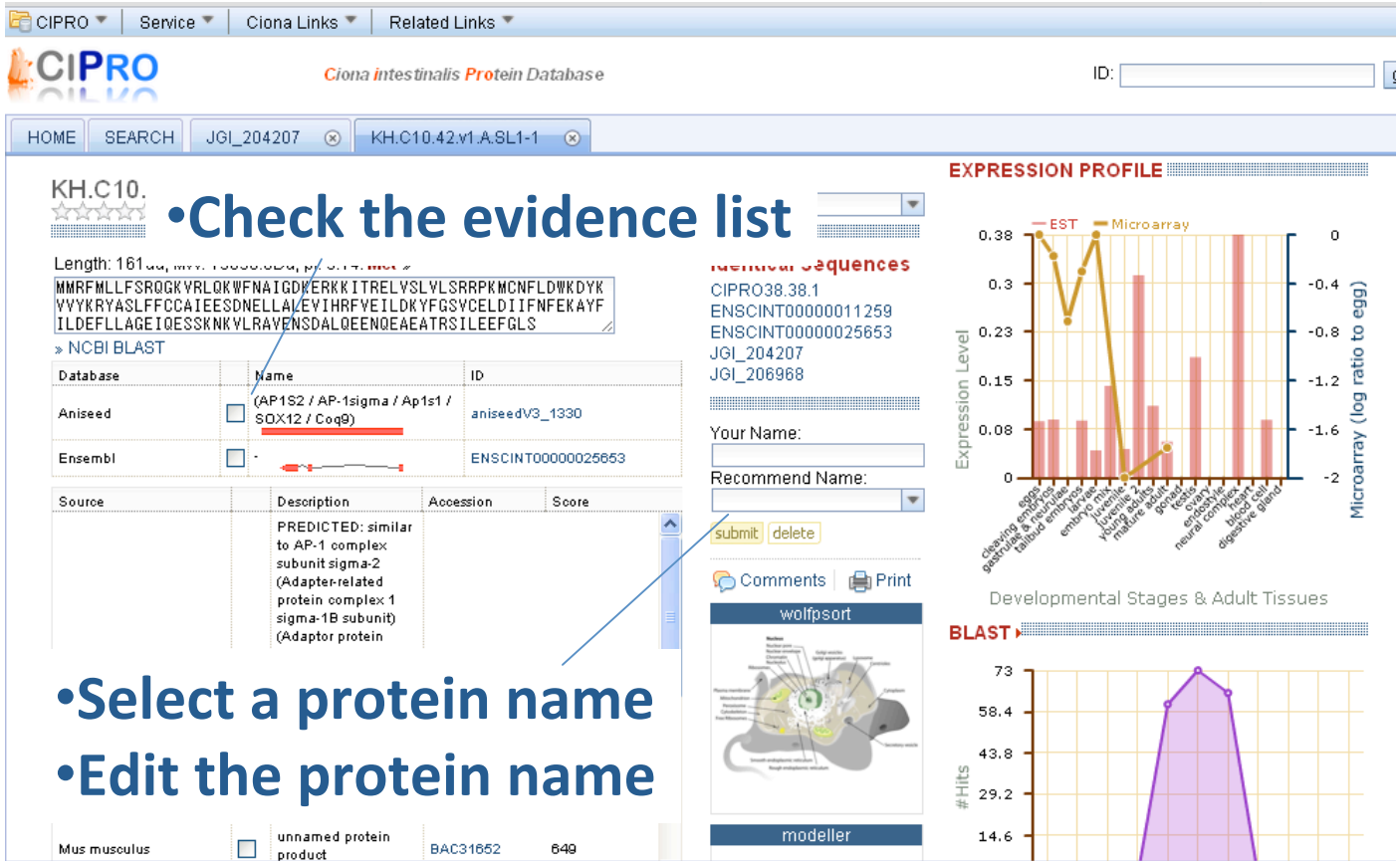
Homology model



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



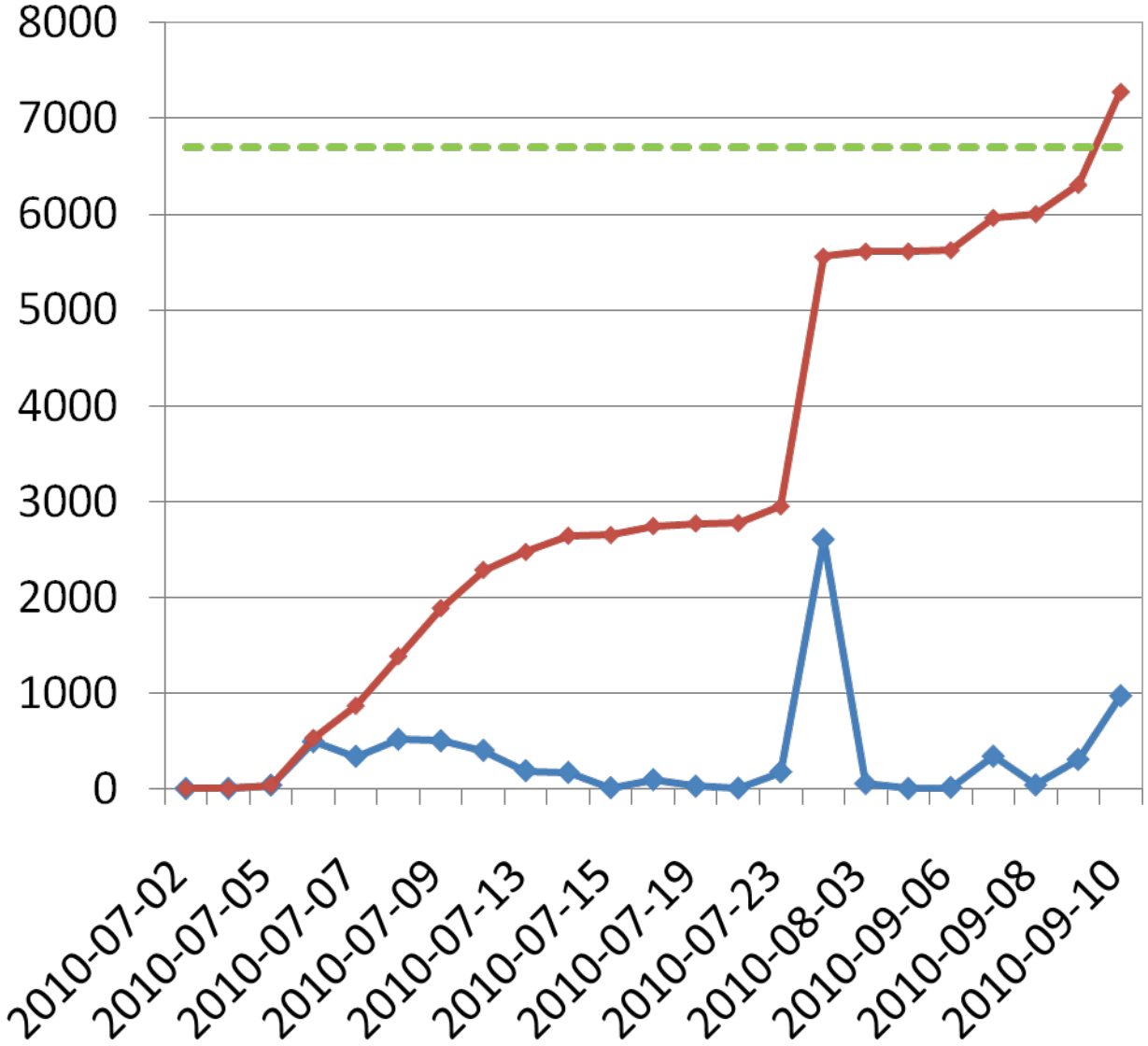
Case study: annotation of protein names by using web forms



How to decide explicit formalism against implicit criteria

- **Class I**: ≥50% identity, ≥ 50% coverage
  - HOMOLOGOUS TO
- **Class II**: ≥ 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

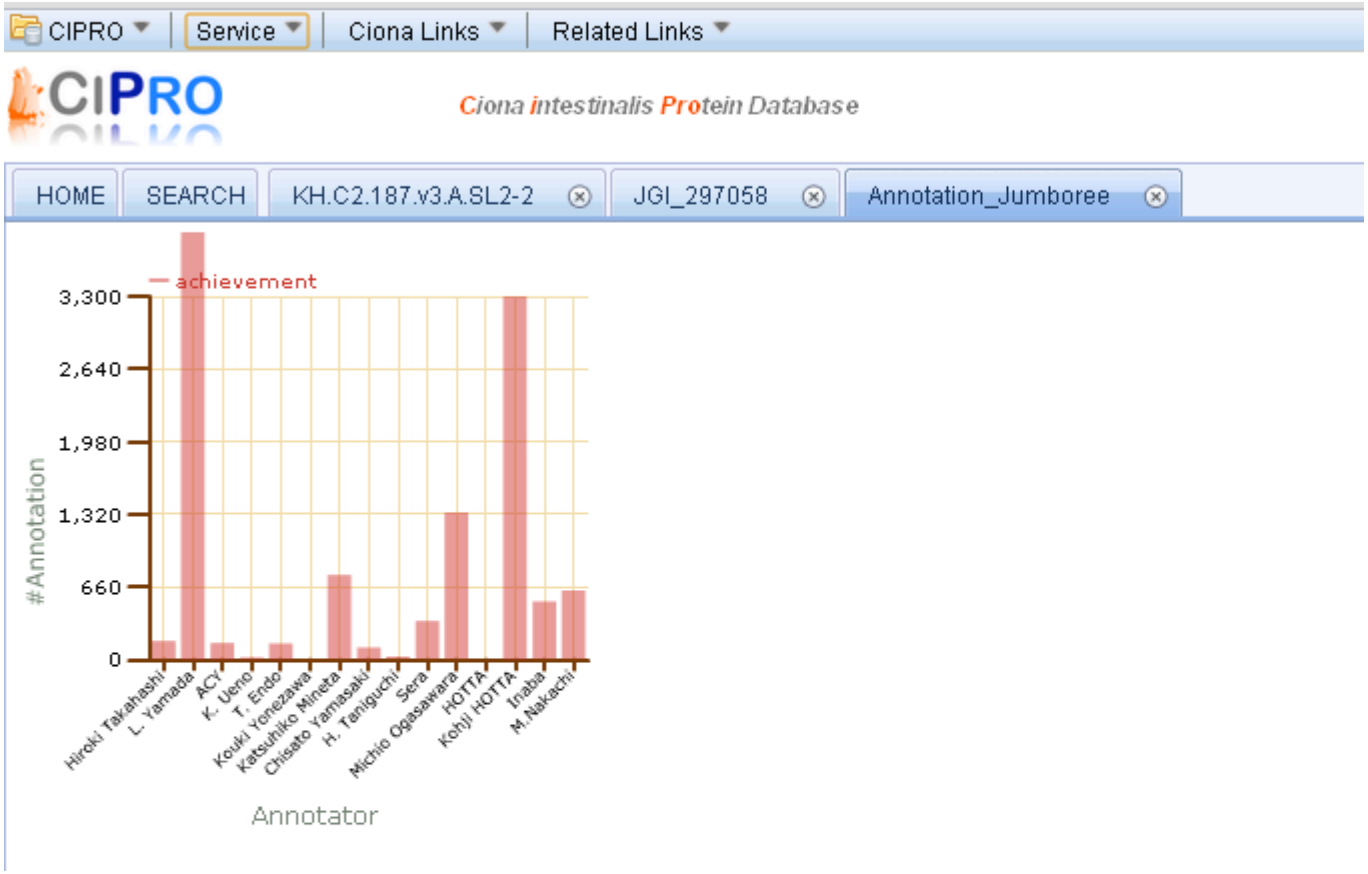
- 1 year ~
- 1 person
- Total #6,697

Flexible descriptions were essentially inescapable

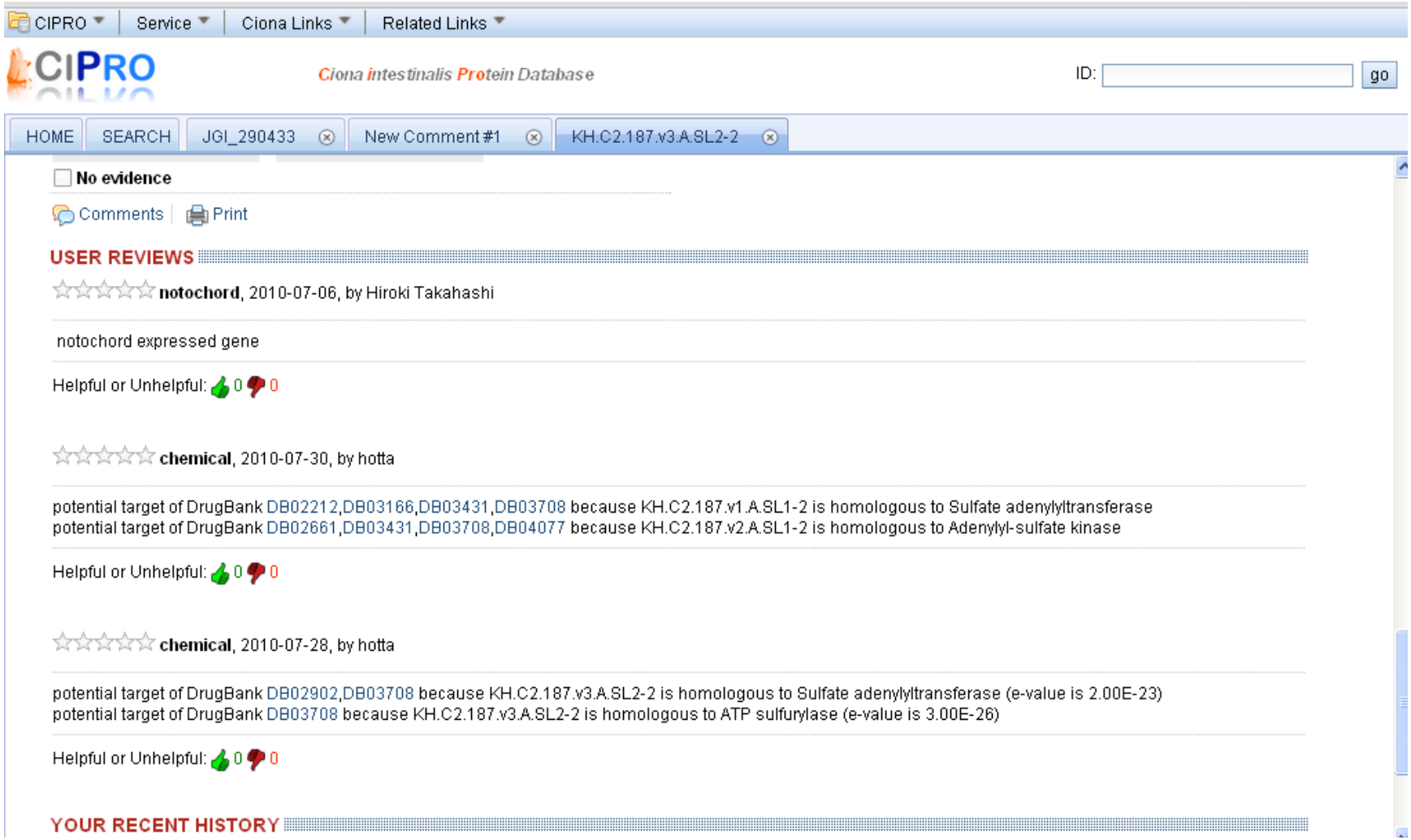
- poly(A)
- poly-A
- polyadenylation

- 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



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).