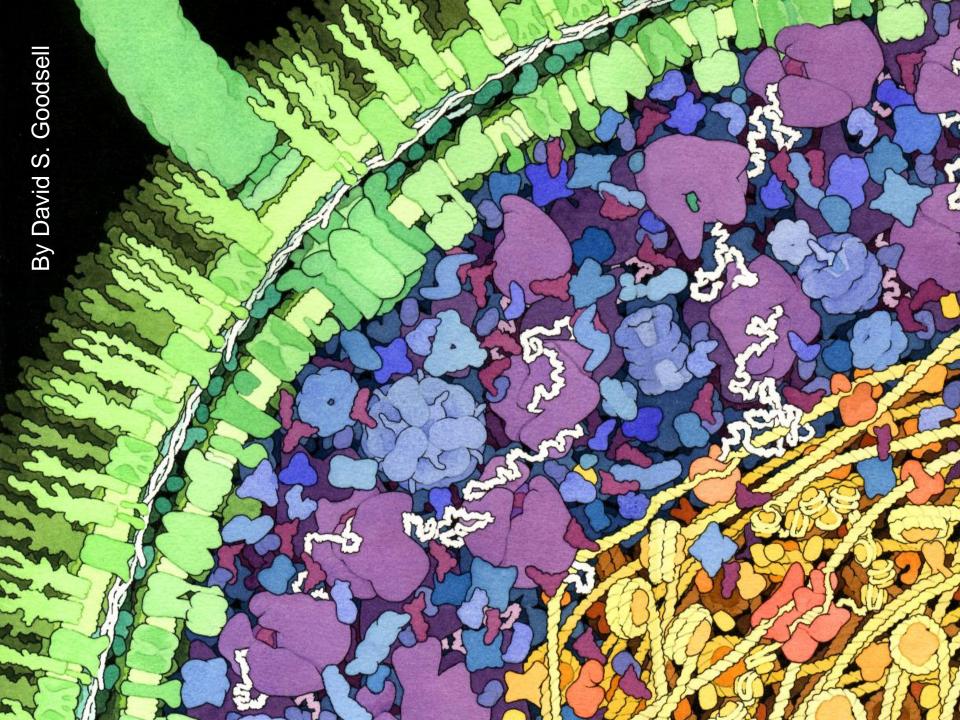
The Pros and Cons of Wiikpedia

Alex Bateman



Anyone can edit anything!



One thousand families for the molecular biologist

Cyrus Chothia

How many families of proteins are there? By putting together the information to be found in papers published over the past few months we can make an initial estimate, and my calculation suggests that the large majority of proteins come from no more than one thousand families.

this work and the proportion of them that have sequences similar to those determined previously.

The six sets of results give a remarkably consistent view: in each case the authors report that close to one-third of the new sequences belong to families that already have members in the sequ-

Nature (1992) 357:543-544.

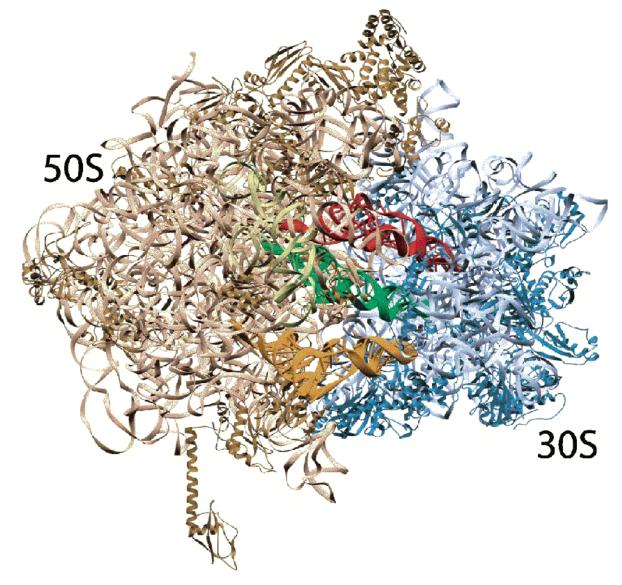
My personal scientific goal:

"to completely and accurately classify all proteins"

The Pfam Protein Families Database



http://pfam.sanger.ac.uk



T. thermophilus - Ramakrishnan, Cell, 2002



A database of non-coding RNA elements

Curating annotation is the bottleneck!



WIKIPEDIA

English

The Free Encyclopedia 1 947 000+ articles

Deutsch

Die freie Enzyklopädie 622 000+ Artikel

Français

L'encyclopédie libre 541 000+ articles

日本語

フリー百科事典 401 000+記事

Nederlands

De vrije encyclopedie 325 000+ artikelen



Polski

Wolna encyklopedia 413 000+ haset

Italiano

L'enciclopedia libera

Português

A enciclopédia livre 279 000+ artigos

Español

La enciclopedia libre 265 000+ artículos

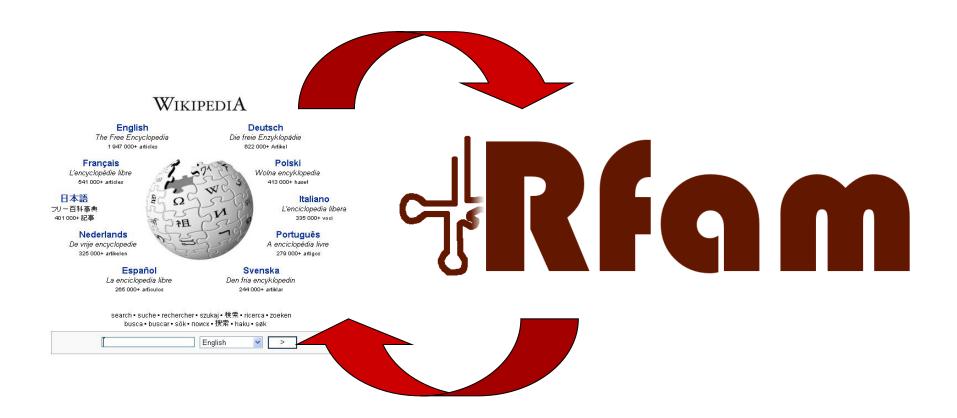
Svenska

Den fria encyklopedin 244 000+ artiklar

search • suche • rechercher • szukaj • 検索 • ricerca • zoeken busca • buscar • sök • поиск • 搜索 • haku • søk

English	>	
English		

The Masterplan

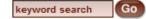


Community Annotation

- We chose to embrace Wikipedia
 - Saves us implementing a wiki
 - Vast audience
 - Low barrier to entry
 - The Google factor







Family: *RyhB* (RF00057)



Summary

Sequences

Alignments

Secondary structure

Species

Trees

Structures

Database references

Curation

Jump to...

enter ID/acc

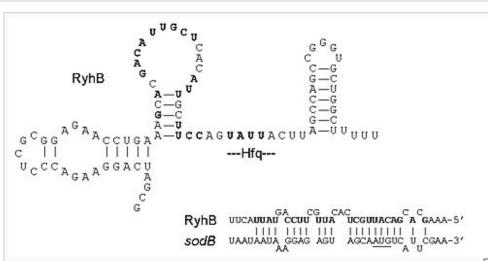


Summary

RyhB RNA Edit Wikipedia Entry

The Rfam group coordinates the annotation of Rfam families in <u>Wikipedia</u> . You can see the Wikipedia page for this family <u>here</u>. <u>More...</u>

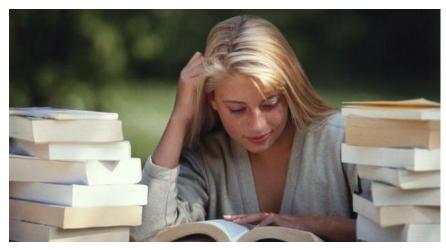
RvhB RNA is a 90 nucleotide noncoding RNA that down-regulates a set of iron-storage and iron-using proteins when iron is limiting; it is itself negatively regulated by the ferric uptake repressor protein, Fur (Ferric uptake regulator). This ncRNA gene was recently identified in a screen and called SraI and was found to be expressed only in stationary phase.[2] RyhB RNA levels are inversely correlated with mRNA levels for the sdhCDAB operon, encoding succinate dehydrogenase, as well as five other genes previously shown to be positively regulated by Fur by an unknown mechanism. These include two other genes encoding enzymes in the tricarboxylic acid cycle, acnA and from A true forcities dended the A and lefe



The figure from Tjaden et al.. depicts the secondary structure for the RyhB RNA. The Sm-like protein Hfq binds to the AU-rich unstructured region of RyhB as indicated. Below the secondary structure, the primary sequence of RyhB is shown along with its putative binding interaction to the target mRNA sodB. The start codon for sodB is underlined. RyhB nucleotides that participate in the interaction are in bold. [1]

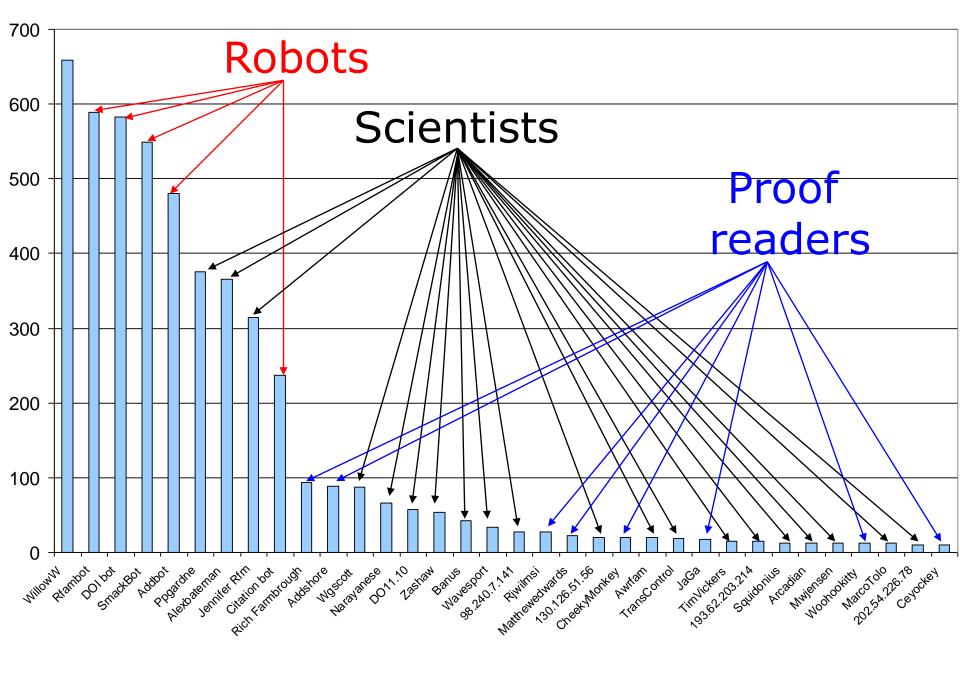
Who is editing wikipedia pages?



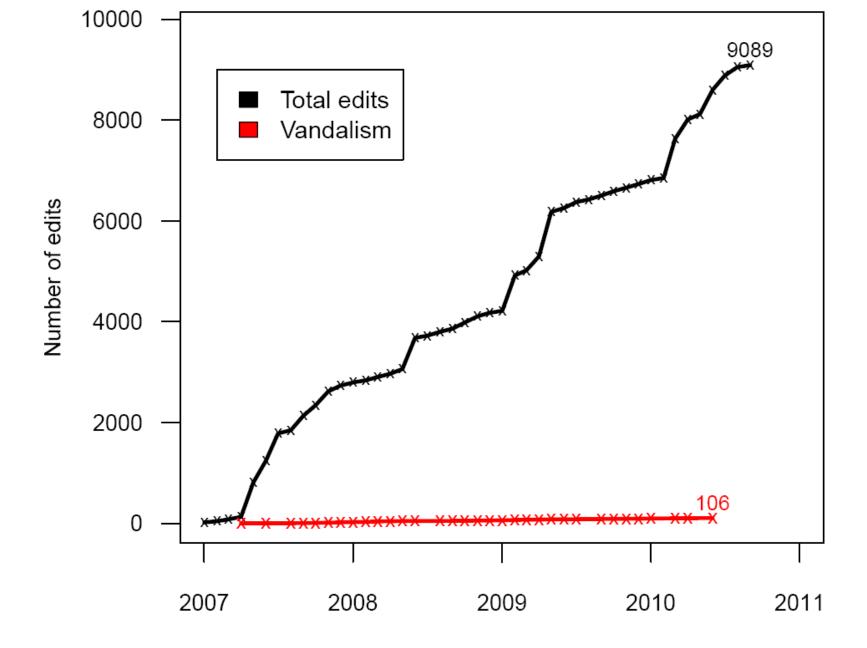








As of 12th April - 1001 users - 7,645 edits



Vandalism



Riboswitches as antibiotic targets

At least one antimicrobial compound, pyrithiamine, has been shown to act by targeting riboswitches. In this case, cells metabolize pyrithiamine to pyrithiamine pyrophosphate, which was shown to bind and activate the TPP riboswitch. As a consequence, the cell ceases to make TPP. It has also been shown that S-(2-aminoethyl)L-cysteine binds the lysine riboswitch; this could explain why this compound is antimicrobial.

One potential advantage that riboswitches have as an antibiotic target is that many of the riboswitches have multiple instances per genome, where each instance controls an operon containing many genes, many of which are essential. Therefore, in order for bacteria to evolve resistance to the antibiotic by mutations in the riboswitch, all riboswitches must be mutated. However, of course, other mechanisms for resistance may exist, and some -- such as altering the specificity of an exporter to export the drug -- may require fewer mutations.

Riboswitches were parodied in an episode of the simpsons where Homer goes on a tour to eat "Ribwiches

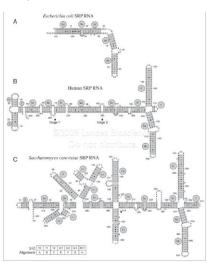
- User <u>203.59.66.54</u> added this line on 3rd June 2007, removed 17th June
- Plausible because Ribwiches did appear in Simpsons in 2003
- Most vandalism is reverted in minutes
- Vandalism only made it onto Rfam website once

Getting scientists to edit

- 10 simple rules article in PLoS CB
 - Help scientists understand how to get started
- December 2008 we struck a deal with RNA Biology.
- RNA families track created
- RNA families articles require:
 - Paper describing family
 - Stockholm formatted alignment
 - Wikipedia article
- Paul Gardner acts as Editor

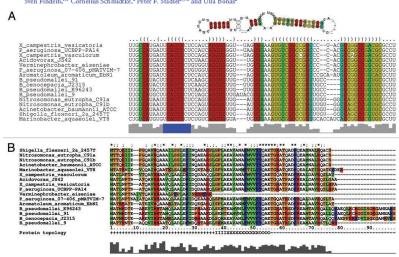
Kinship in the SRP RNA family

Magnus Alm Rosenblad,1 Niels Larsen,2 Tore Samuelsson3 and Christian Zwieb4.*



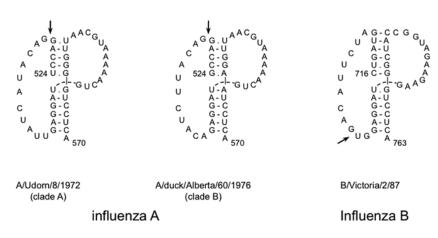
A novel family of plasmid-transferred anti-sense ncRNAs

Sven Findeiß,1,* Cornelius Schmidtke,2 Peter F. Stadler1,3-6 and Ulla Bonas2

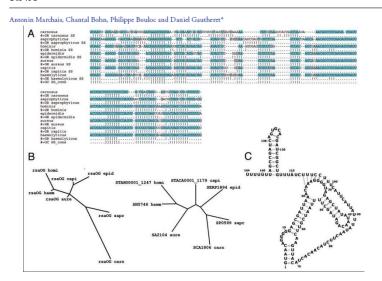


A family of non-classical pseudoknots in influenza A and B viruses

Alexander P. Gultyaev^{1,2*} and René C.L. Olsthoorn¹



RsaOG, a new staphylococcal family of highly transcribed non-coding RNA



The problem with scientists

Ignorant Scientists Knowledge of subject Wikipedia **Pedants**

Knowledge of Wikipedia

Vanity

- Most scientists gravitate to writing articles about themselves, their groups or projects
- This is an obvious conflict of interest and is discouraged by Wikipedia
- However, writing about a subject you are an expert in is not!

 http://en.wikipedia.org/wiki/Wikipedia:An_article_ about_yourself_is_nothing_to_be_proud_of

The ChEMBL group is a bioinformatics and cheminformatics group based at the European Bioinformatics Institute (EBI), Hinxton, Cambridgeshire. ChEMBL interests cover Computational Chemical Biology, Chemogenomics, Chemoinformatics, Bioinformatics, Structural Biology, Open Data, Knowledge Management, and Data Integration.

ChEMBLdb

ChEMBLdb @ is a database of bioactive drug-like small molecules, it contains 2-D structures, calculated properties (e.g. logP, Molecular Weight, Lipinski Parameters, etc.) and abstracted bioactivities (e.g. binding constants, pharmacology and ADMET data).

Data Information

The bioactivities are normalised into a uniform set of end-points and units, where possible, and the links between a molecular target and a published assay are tagged with a set of varying confidence levels. The data is abstracted and curated from primary scientific literature and covers a significant fraction of the Structure-Activity Relationships (SAR) and discovery of modern drugs.

At the time of the release of ChEMBL_07 (Oct 2010), ChEMBLdb contained data on 8,078 targets, 602,500 distinct compounds (730,186 compound records) and 2,948,069 activities from 38,204 publications.

Additional data on the clinical progress of compounds is being integrated into ChEMBL at the current time.

Other ChEMBL Databases

Subsets of ChEMBLdb, relating to particular target classes (Kinase SARfari &, GPCR SARfari &), or disease areas (ChEMBL-NTD &), are also available.

These separate data sets, and the entire ChEMBLdb, are available either via download dthrough the website, or via bespoke query interfaces, tailored to the requirements of the scientific communities with a specific interest in these research areas.

ChEMBL Blog

The ChEMBL group keeps their users up to date via their group blog - ChEMBL Blog . The blog informs users about releases, new entries, group activities and future plans. It is also a good place to find out about web walk-throughs and testing of new interfaces.

Deletion log

(del/undel) 21:28, 8 October 2010 Alexf (talk | contribs | block) deleted "ChEMBL" (G11: Unambiguous advertising or promotion) (view/restore)

Restore revisions

To restore the entire page and its history, leave all checkboxes deselected and click **Restore**. To perform a selective restoration, check the boxes corresponding to the revisions to be restored and click **Restore**. Selecting a box, then shift selecting another will fill all boxes in between in many browsers. Clicking **Reset** will clear the comment field and all checkboxes. Please make sure that you are following undeletion policy and that you leave a summary in the comment box.

Comment: | Invert selection | Restore | Reset | Select all

Page history

- □ (del/undel) (diff) 20:27, 8 October 2010 . . Jimmy Pitt (talk | contribs | block) (2,612 bytes) (Requesting speedy deletion (CSD G11). (TW))
- ☐ (del/undel) (diff) 11:46, 8 October 2010 . . Chemblgroup (talk | contribs | block) (2,600 bytes)
- (del/undel) (diff) 11:43, 8 October 2010 . . Chemblgroup (talk | contribs | block) (2,404 bytes)
- (del/undel) (diff) 11:42, 8 October 2010 . . Chemblgroup (talk | contribs | block) (2,411 bytes)
- (del/undel) (diff) 11:40, 8 October 2010 . . Chemblgroup (talk | contribs | block) (2,376 bytes)
- (del/undel) (diff) 11:36, 8 October 2010 . . Chemblgroup (talk | contribs | block) (2,363 bytes)
- (del/undel) (diff) 11:19, 8 October 2010 . . Chemblgroup (talk | contribs | block) (2,395 bytes)
- □ (del/undel) (diff) 11:18, 8 October 2010 . . Chemblgroup (talk | contribs | block) (2,367 bytes)
- ☐ (del/undel) (diff) 11:17, 8 October 2010 . . Chemblgroup (talk | contribs | block) (2,320 bytes) (←Created page with ""ChEMBL" is a bioinformatics and cheminformatics group based at the European Bioinformatics Institute (EBI), Hinxton, Camb...")

discussion

edit this page new section user

page

User talk:Chemblgroup

From Wikipedia, the free encyclopedia



Welcome to Wikipedia. Because we have a policy against usernames that give the impression that the account represents a group, organization or website, I have blocked this account; you are welcome to create a new account with a username that represents only you. You should also read our conflict of interest guideline. If your username doesn't represent a group, organization or website, you may appeal this username block by adding below this notice the text {{unblock|Your reason here}}. Thank you, Wifione Leave a message 12:45 pm, 8 October 2010, Friday (1 month, 19 days ago) (UTC+1)

Speedy deletion nomination of ChEMBL

[edit]



A tag has been placed on ChEMBL, requesting that it be speedily deleted from Wikipedia. This has been done under section G11 of the criteria for speedy deletion, because the page seems to be unambiguous advertising which only promotes a company, product, group, service or person and would need to be fundamentally rewritten in order to become an encyclopedia article. Please read the guidelines on spam as well as Wikipedia:FAQ/Business for more information. You may also wish to consider using a Wizard to help you create articles - see the Article Wizard.

If you think that this notice was placed here in error, you may contest the deletion by adding {{hangon}} to the top of the page that has been nominated for deletion (just below the existing speedy deletion or "db" tag - if no such tag exists then the page is no longer a speedy delete candidate and adding a hangon tag is unnecessary), coupled with adding a note on the talk page explaining your position, but be aware that once tagged for speedy deletion, if the page meets the criterion, it may be deleted without delay. Please do not remove the speedy deletion tag yourself, but don't hesitate to add information to the page that would render it more in conformance with Wikipedia's policies and guidelines. Lastly, please note that if the page does get deleted, you can contact one of these administrators to request that they userfy the page or have a copy emailed to you. Jimmy Pitt talk 8:27 pm, 8 October 2010, Friday (1 month, 19 days ago) (UTC+1)

Categories: Wikipedians who are indefinitely blocked for promotional user names

ChEMBL

From Wikipedia, the free encyclopedia

ChEMBL or **ChEMBLdb** is a manually curated chemical database of bioactive molecules with drug-like properties. ^[1] It is maintained by the European Bioinformatics Institute (EBI), based on the Wellcome Trust Genome Campus, Hinxton, UK. The database, orginally known as StARlite, was developed by a pharmaceutical company, Galapagos NV. It was was acquired for EMBL in 2008 with an award from The Wellcome Trust, resulting in the creation of the ChEMBL chemogenomics group at EBI, led by John Overington. ^{[1][2]}



Contents [show]

Scope and access [edit]

ChEMBL version 2 was launched in January 2010, including data about 2.4 million activities of 622,824 compounds, including 24,000 natural products. This was obtained from nearly 34,000 publications across twelve medicinal chemistry journals. According to Andreas Bender writing in *Nature Chemical Biology*, ChEMBL's coverage of available bioactivity data is "the most comprehensive ever seen in a public database." In October 2010 ChEMBL version 8 was launched, with over 2.97 million activities of 636,269 compounds.

ChEMBLdb can be accessed via a web interface or downloaded by File Transfer Protocol. It is formated in a manner amenable to computerized data mining, and attempts to standardize activities between different publications, to enable comparative analysis. [1] ChEMBL is also integrated into other large-scale chemistry resources, including PubChem and the ChemSpider system of the Royal Society of Chemistry.

Associated resources [edit]

In addition to the database, the ChEMBL group have developed tools and resources for data mining. These include Kinase SARfari, an integrated chemogenomics workbench focussed on kinases. The system incorporates and links sequence, structure, compounds and screening data. GPCR SARfari is a similar workbench focussed on GPCRs and ChEMBL-Neglected Tropical Disease (ChEMBL-NTD) is a repository for Open Access primary screening and medicinal chemistry data directed at endemic tropical diseases of the developing regions of the Africa, Asia, and the Americas. The primary purpose of ChEMBL-NTD is to provide a freely accessible and permanent archive and distribution centre for deposited data. [1]

See also

- ChEBI
- DrugBank

References [edit]

- 1. A a b c d e f Bender, A (2010). "Databases: Compound bioactivities go public" . Nature Chemical Biology 309 (6). doi:10.1038/nchembio.354 . Retrieved 11-15-2010.
- 2. * Overington J (April 2009). "ChEMBL. An interview with John Overington, team leader, chemogenomics at the European Bioinformatics Institute Outstation of the European Molecular Biology Laboratory (EMBL-EBI). Interview by Wendy A. Wart". J. Comput. Aided Mol. Des. 23 (4): 195–8. doi:10.1007/s10822-009-9260-9 . PMID 19194660 .
- 3. ^ ChEMBL-og (15 November 2010), ChEMBL_08 Released 🗗, retrieved 11-15-2010

External links [edit]

- ChEMBLdb.
- Kinase SARfari
- ChEMBL-Neglected Tropical Disease Archive
- GPCR SARfari
- The ChEMBL-og

 Open data and drug discovery blog run by the ChEMBL team.

Categories: Bioinformatics | Biological databases | Chemical databases | Cheminformatics | Wellcome Trust

User Alexm1313 create page on Frozen

Ark on 19th July.

Page rapidly grows



The Frozen Ark Project is a frozen zoo created jointly by the Zoological Society of London, the Natural History Museum, and Nottingham University¹⁰. The Frozen Ark project aims to preserve the DNA of endangered species to retain the genetic knowledge for our future. The Frozen Ark collects and stores DNA samples are taken from animals in zoos and those threatened with extinction in the wild. Currently, animals are facing a huge amount of extinction, and despite great efforts by conservationists, this looks set to continue. When these animals die out, their genetic material is lost. The Frozen Ark project retains that genetic material which will be very important in the future of science. It is expected that over the next 30 years, over a quarter of the world's animals will become extinct. The destruction of the environments of many invertebrates are leading to hug losses before they have even been characterised by scientists. This project shares similarity with the Kew Gardens Millennium Seed Bank Project, that aims to preserve all the world's seeds. The genetic information from the preserved DNA will lead on to key scientists as a compendium of knowledge about the species biology, behaviour and evolution.



Aims of the Frozen Ark

- Create and continue promoting centres that will aid conservation of endangered animal DNA.
- · Create expert groups in collecting material from different animals.
- . Construct a database of the species DNA currently held in the Frozen Ark storage world-wide.
- . Animals with greatest risk of extinction are given high priority over others.
- Set up expert panel groups for each animal group.
- Provide resources for the scientific community to aid research.
- Provide material to help stop the loss of diversity and aid conservation programmes world-wide [2].

The Frozen Ark Consortium

The Frozen Ark Consortium includes world-wide institutions that are all making an effort to continue the aims of the Frozen Ark project. The charity is based at and organised from The Frozen Ark Unit at the School of Biology at the University of Nottingham, United Kingdom. It is also here where the main laboratory is based where the DNA is frozen and stored long-term.

The Frozen Ark consortium consists of the following groups:

Content dispute [edit]

I have tried to engage Alexm1313 (talk • contribs) in a discussion about content he has added to this article, specifically material in the lede regarding the prospects of future extinctions and the supposed efficacy of this project. As I pointed out on his talk page:

- There is no need for a discussion of rates of extinction on this page, as that topic is thoroughly covered elsewhere in Wikipedia, and covering it
 here simply means that the information must be maintained in multiple places.
- There is no evidence that the information will be extremely important to future scientists. Clearly, there is a hope that this will be the case (else
 the project is useless), but there is no evidence to support this claim.
- Stylistically, one would never say "animals are facing a huge amount of extinction".
- The expectation that 1/4 of the world's species will go extinct over the next 30 years is uncited, and is unnecessary in this article. By linking to the
 article on extinction, we can provide a link to the necessary information without duplicating it (or providing incorrect information) in this article.
 Duplication of information across multiple articles is called content forking and is discouraged.

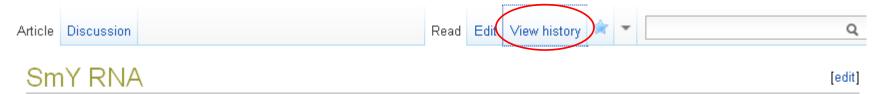
However I have been unable to elicit a response from this user other than a continued re-insertion of the disputed material. Other opinions are welcome. WikiDan61 ChatMe! ReadMe!! 17:28, 19 July 2010 (UTC)

Lots of discussions later ...

I am not interested in people being horrible, and I want no affiliation to that page. Poor newbies who have to put up with that. Wasted hours of my work time for something very simple. That is it for me, I am not interested in that kind of behaviour.

Attribution

Many scientists are worried about lack of attribution

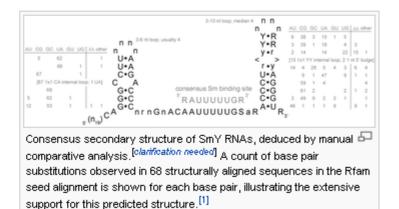


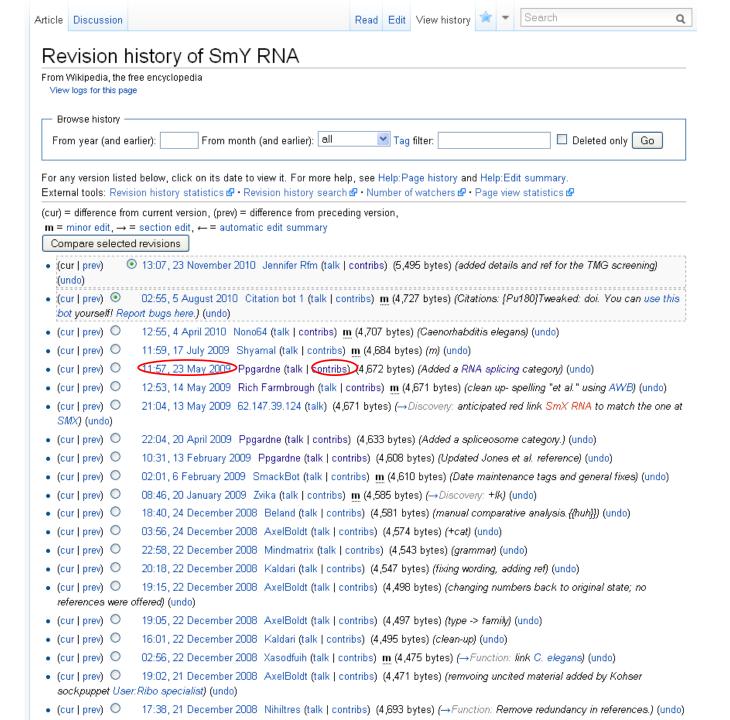
A C-class article from Wikipedia, the free encyclopedia

SmY ribonucleic acids (SmY RNAs) are a family of small nuclear RNAs found in some species of nematode worms. They are thought to be involved in mRNA trans-splicing.

SmY RNAs are about 70-90 nucleotides long and share a common secondary structure, with two stem-loops flanking a consensus binding site for Sm protein. [2][3] Sm protein is a shared component of spliceosomal snRNPs.

SmY RNAs have been found in nematodes of class Chromadorea, which includes the most commonly studied nematodes (such as *Caenorhabditis*, *Pristionchus*, and *Ascaris*), but not in the more distantly related *Trichinella spiralis* in class Dorylaimia. The number of SmY genes in each species varies, with most *Caenorhabditis* and *Pristionchus* species having 10-26 related paralogous copies, while other nematodes have 1-5.^[1]





- 10:30, 26 November 2010 (diff | hist) Plasmodium falciparum (Undid revision 398936168 by 82.215.36.66 (talk)do you have a citation for that?)
 (top)
- 14:39, 25 November 2010 (diff | hist) MiR-122 (Sentence belongs with paragraph above) (top)
- 19:42, 24 November 2010 (diff | hist) Nm User talk: Squabblefish (Added welcome template to user talk page using TW) (top)
- 15:46, 24 November 2010 (diff | hist) Mir-132 (→Further reading: another ref.) (top)
- 13:59, 24 November 2010 (diff | hist) History of RNA biology (+Category:Non-coding RNA; +Category:History of science using HotCat)
- 13:58, 24 November 2010 (diff | hist) History of RNA biology (added Category: RNA using HotCat).
- 13:37, 24 November 2010 (diff | hist) Coronavirus 3' stem-loop II-like motif (s2m) (more better englisher.) (top)
- 11:10, 24 November 2010 (diff | hist) Nm User talk:83.249.23.21 (Added welcome template to user talk page using TW) (top)
- 19:52, 23 November 2010 (diff | hist) Coronavirus 3' stem-loop II-like motif (s2m) (added a tertiary structure image)
- 17:02, 23 November 2010 (diff | hist) Cell division (Undid revision 398451008 by 81.105.171.111 (talk)) (top)
- 13:57, 23 November 2010 (diff | hist) List of biological databases (+Category: Biological databases; +Category: Bioinformatics using HotCat) (top)
- 13:24, 23 November 2010 (diff | hist) Talk:List of RNA structure prediction software (added to list class) (top)
- 13:23, 23 November 2010 (diff | hist) Nm User talk: Michipanero (Added welcome template to user talk page using TW) (top).
- 12:34, 23 November 2010 (diff | hist) Talk: Biological database (→Link list is poor article style: typo) (top)
- 12:33, 23 November 2010 (diff | hist) Talk: Biological database (→Link list is poor article style: moving).
- 12:31, 23 November 2010 (diff | hist) User talk: Ppgardne (→WTSI wiki-group) (top)
- 12:28, 23 November 2010 (diff | hist) N Talk: List of biological databases (mcb template)
- 12:27, 23 November 2010 (diff | hist) m List of biological databases (moved List of Biological databases to List of biological databases).
- 12:27, 23 November 2010 (diff | hist) N List of Biological databases (moved List of Biological databases to List of biological databases) (top)
- 12:26, 23 November 2010 (diff | hist) List of biological databases (reference section)
- 12:25, 23 November 2010 (diff | hist) N List of biological databases (created page)
- 12:22, 23 November 2010 (diff | hist) Nm User talk: Magdalena ZZ (Added welcome template to user talk page using TW) (top).
- 06:05, 23 November 2010 (diff | hist) Alan MacDiarmid (Undid revision 398375017 by 137.28.55.61 (talk))
- 11:37, 22 November 2010 (diff | hist) N Talk: Efflux (microbiology) (added MCB template) (top)
- 19:08, 21 November 2010 (diff | hist) Nm User talk:83.94.219.205 (Added welcome template to user talk page using TW) (top).
- 19:07, 21 November 2010 (diff | hist) Nm User talk: Sgj67 (Added welcome template to user talk page using TW) (top)
- 21:42, 20 November 2010 (diff | hist) Mir-1 microRNA precursor family (added a further reading section & cleaned up some mal-formed refs.) (top)
- 20:14, 20 November 2010 (diff | hist) Let-7 microRNA precursor (added a further reading section) (top)
- 20:00, 20 November 2010 (diff | hist) Mir-17 microRNA precursor family (added a further reading section) (top)

Usemame: <u>Ppgardne</u> Apr 08, 2006 12:52:12 First edit: 768 Unique pages edited: Average edits per page: 2.67 2,033 Live edits: 17 Deleted edits: Total edits (including deleted): 2,050 Namespace Totals Namespace Totals 1560 76.73% 91 4.48% User 159 7.82% -4.48 82 4.03% User talk Article 37 1.82% Talk Wikipedia User Wikipedia talk 14 0.69% Usertalk 54 2.66% Wikipedia Wikipedia talk 0.10% File Template 13 0.64% Template 76.73-Template talk 15 0.74% Template talk 0.20% Portal 0.10% Month counts 2006/04 5 2006/06 0 2006/06 0 2006/07 0 2006/08 0 2006/09 2 2006/10 O 2006/11 0 2006/12 0 2007/01 0 2007/02 0 2007/03 46 2007/04 0 2007/05 6 2007/06 16 2007/07 182 2007/08 4 2007/09 6 2007/10 13 2007/11 9 2007/12 12 2008/01 35 2008/02 5 2008/03 13 2008/04 1 2008/06 25 2008/06 10 2008/07 6 2008/08 14 2008/09 6 2008/10 27 2008/11 109 2008/12 9 2009/01 0 2009/02 35 2009/03 37 2009/04 108 2009/06 150 2009/06 39 2009/07 19 2009/08 27 2009/09 55 2009/10 52 2009/11 28 2009/12 50 20 10/0 1 24 20 10/02 29

Attribution

- Scientists still want to know what do I get out of editing Wikipedia
- Do tenure committees consider this, etc...

This rather misses the point

