Tin Sinh học
Bioinformatics

Bài thực hành 1. Hướng dẫn sử dụng NCBI

Tài liệu tham khảo

Nicholas James Provart, Bioinformatic Methods I, Coursera, University of Toronto, 2021.

2 bài Lab

- Lab 1a. Exploring NCBI
- Lab 1b. Basic BLAST

Nội dung báo cáo

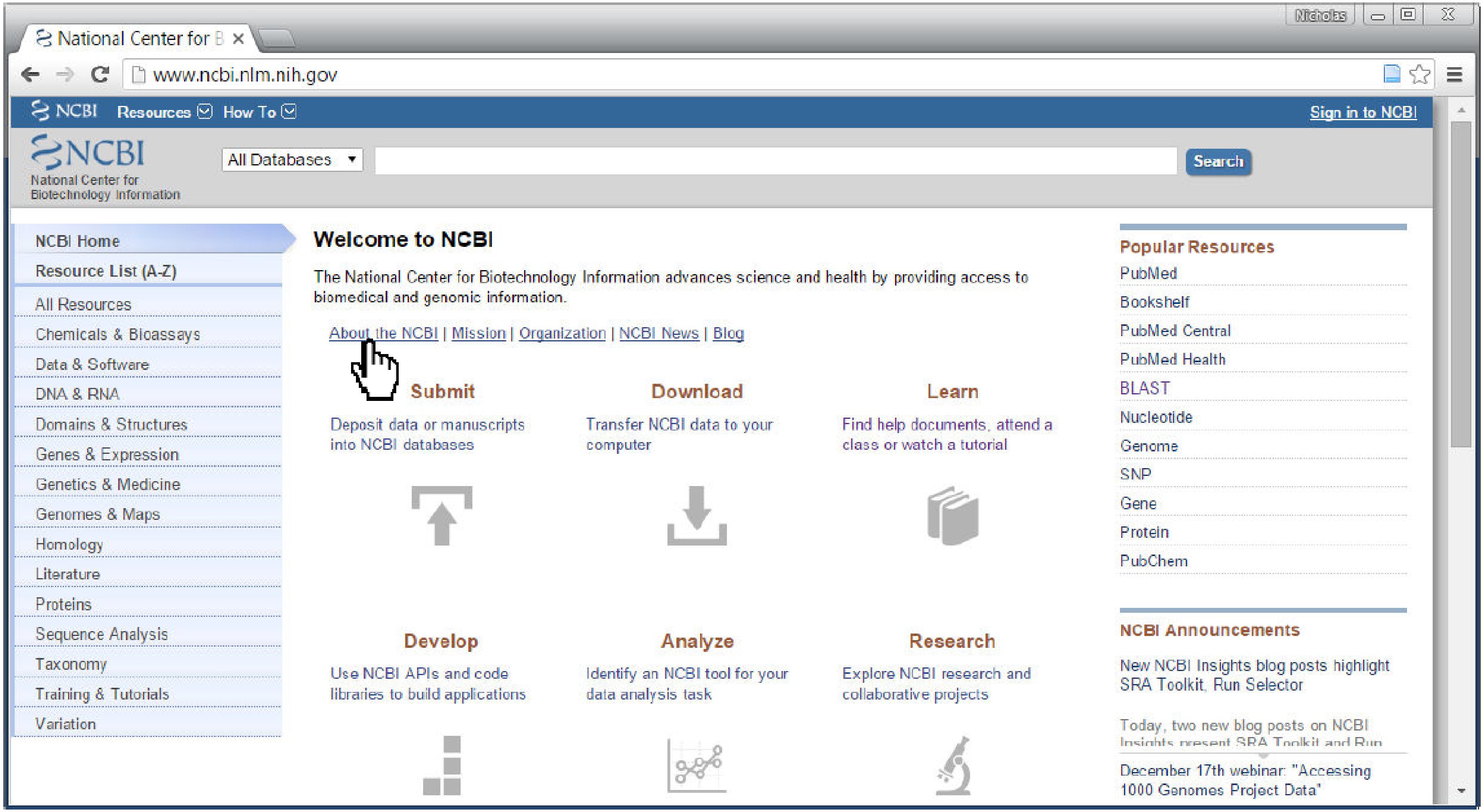
- Mô tả các bước thực hành.
- Trả lời các câu hỏi.
- Chụp màn hình kết quả thu được.
- Download các file trình tự.
- Nộp các kết quả trên vào Assignment trên Microsoft Team.

Lab 1a. Exploring NCBI

Bước 1.

- Open your Web browser and go to NCBI’s homepage: www.ncbi.nlm.nih.gov.
- This page provides links to all of NCBI databases and resources.
- Click ***About the NCBI*** to go a page summarizing some of these resources.

The NCBI homepage



Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

Submit

Deposit data or manuscripts into NCBI databases

Download

Transfer NCBI data to your computer

Learn

Find help documents, attend a class or watch a tutorial

Develop

Use NCBI APIs and code libraries to build applications

Analyze

Identify an NCBI tool for your data analysis task

Research

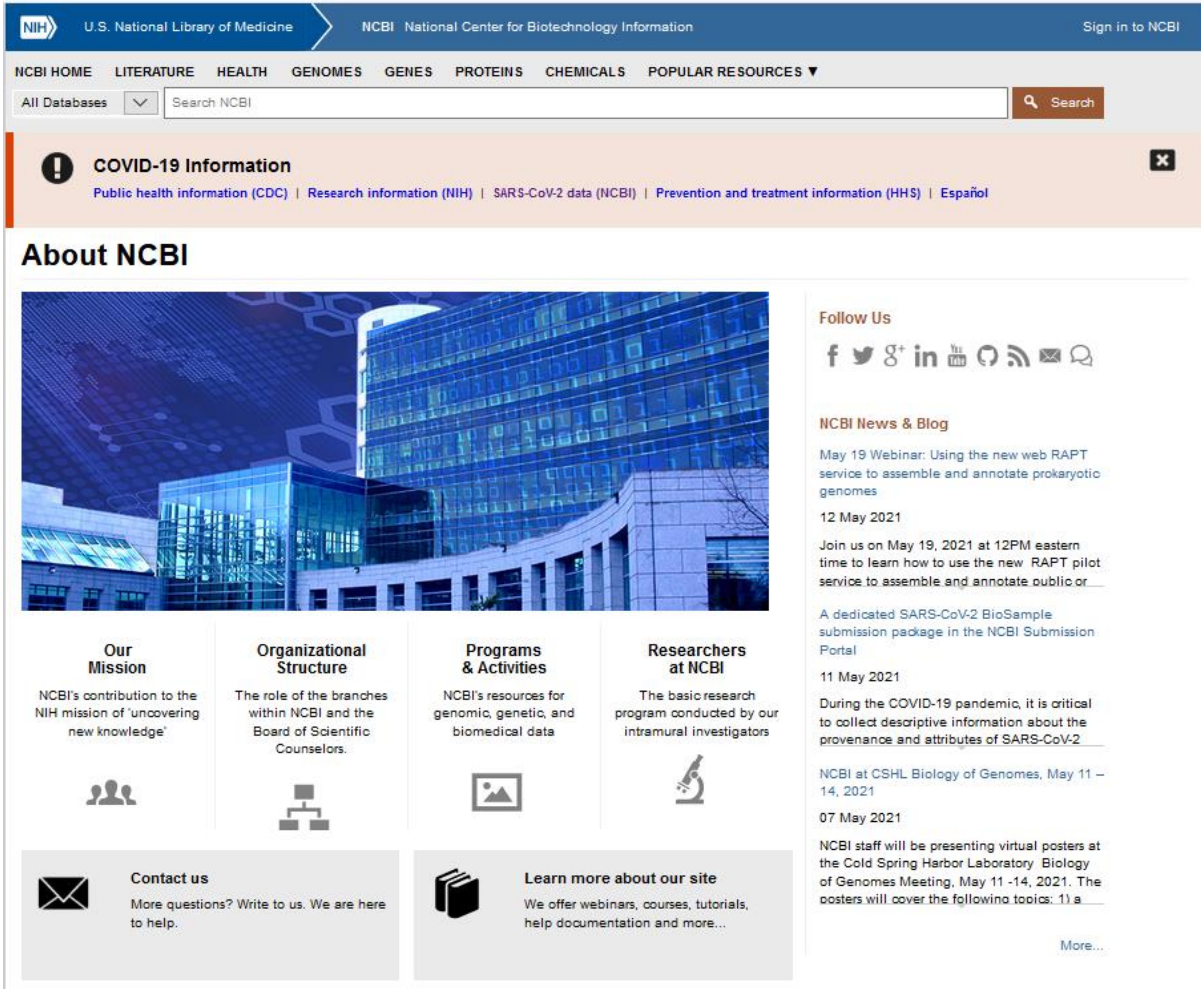
Explore NCBI research and collaborative projects

NCBI Announcements

New NCBI Insights blog posts highlight SRA Toolkit, Run Selector

Today, two new blog posts on NCBI Insights present SRA Toolkit and Run

December 17th webinar: "Accessing 1000 Genomes Project Data"



COVID-19 Information

Public health information (CDC) | Research information (NIH) | SARS-CoV-2 data (NCBI) | Prevention and treatment information (HHS) | Español

About NCBI

Our Mission

NCBI's contribution to the NIH mission of 'uncovering new knowledge'

Organizational Structure

The role of the branches within NCBI and the Board of Scientific Counselors.

Programs & Activities

NCBI's resources for genomic, genetic, and biomedical data

Researchers at NCBI

The basic research program conducted by our intramural investigators

Contact us

More questions? Write to us. We are here to help.

Learn more about our site

We offer webinars, courses, tutorials, help documentation and more...

Follow Us

NCBI News & Blog

May 19 Webinar: Using the new web RAPT service to assemble and annotate prokaryotic genomes

12 May 2021

Join us on May 19, 2021 at 12PM eastern time to learn how to use the new RAPT pilot service to assemble and annotate public or

A dedicated SARS-CoV-2 BioSample submission package in the NCBI Submission Portal

11 May 2021

During the COVID-19 pandemic, it is critical to collect descriptive information about the provenance and attributes of SARS-CoV-2

NCBI at CSHL Biology of Genomes, May 11 – 14, 2021

07 May 2021

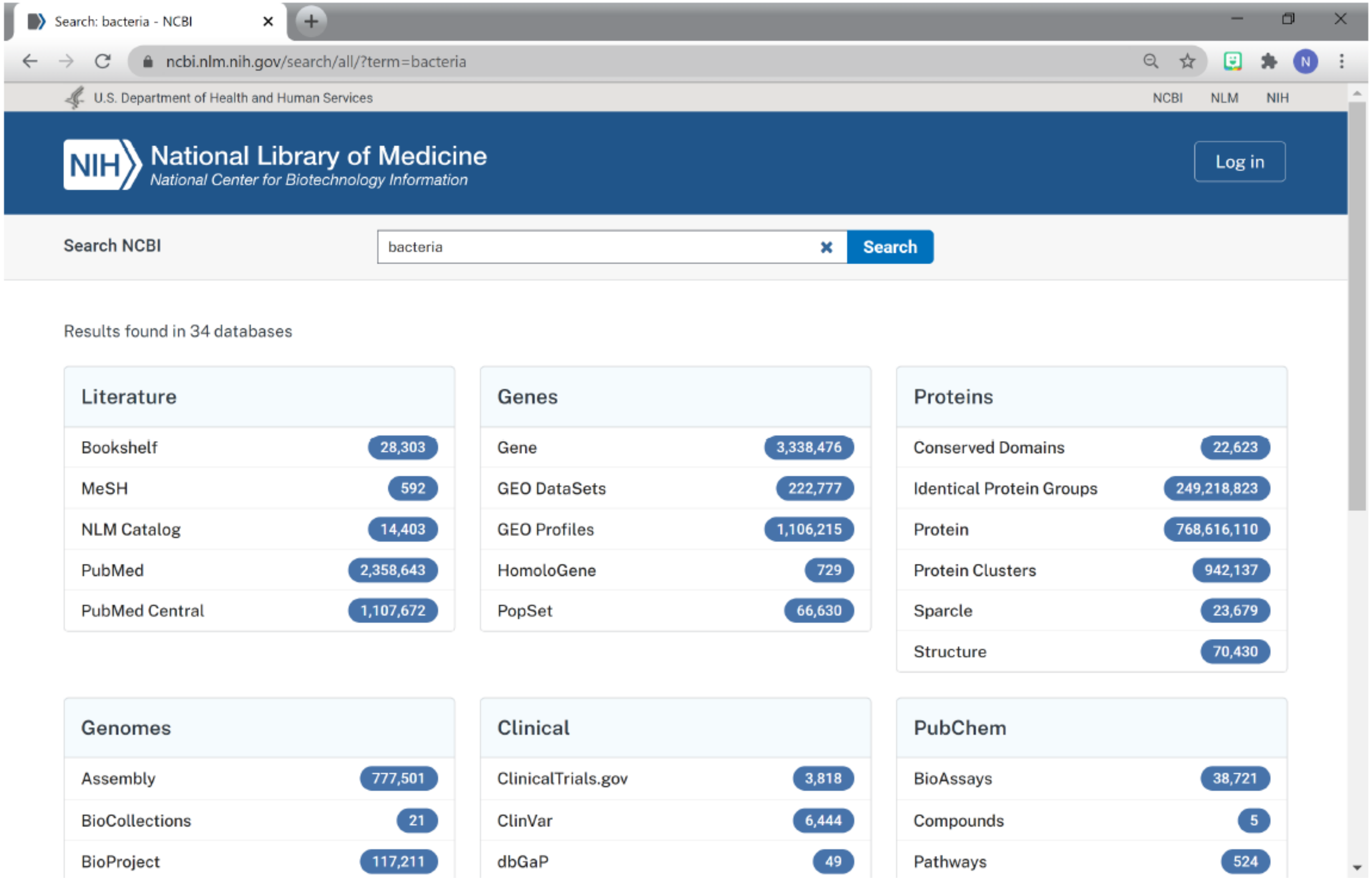
NCBI staff will be presenting virtual posters at the Cold Spring Harbor Laboratory Biology of Genomes Meeting, May 11 -14, 2021. The posters will cover the following topics: 1) a

More...

Bước 2.

- Move to the Search NCBI, select All Databases from the navigation bar at the top of the NCBI start page and click “Search” beside the empty field.
- In the "Search NCBI" box, type in bacteria.
- The output is a summary page of the number of hits in each section.

The Search NCBI portal page with bacteria used as a search word



Search: bacteria - NCBI
ncbi.nlm.nih.gov/search/all/?term=bacteria
U.S. Department of Health and Human Services
NCBI
NLM
NIH
NIH
National Library of Medicine
National Center for Biotechnology Information
Log in
Search NCBI
bacteria
Search

Results found in 34 databases

Literature	
Bookshelf	28,303
MeSH	592
NLM Catalog	14,403
PubMed	2,358,643
PubMed Central	1,107,672

Genes	
Gene	3,338,476
GEO DataSets	222,777
GEO Profiles	1,106,215
HomoloGene	729
PopSet	66,630

Proteins	
Conserved Domains	22,623
Identical Protein Groups	249,218,823
Protein	768,616,110
Protein Clusters	942,137
Sparcle	23,679
Structure	70,430

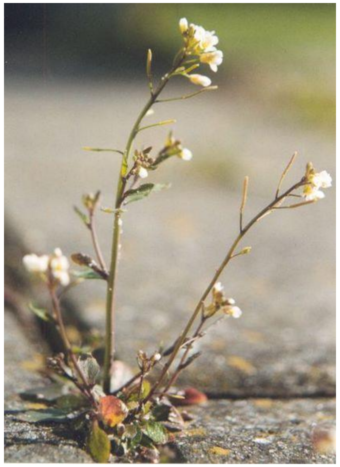
Genomes	
Assembly	777,501
BioCollections	21
BioProject	117,211

Clinical	
ClinicalTrials.gov	3,818
ClinVar	6,444
dbGaP	49

PubChem	
BioAssays	38,721
Compounds	5
Pathways	524

Bước 3. Chọn protein *NP_001318308*

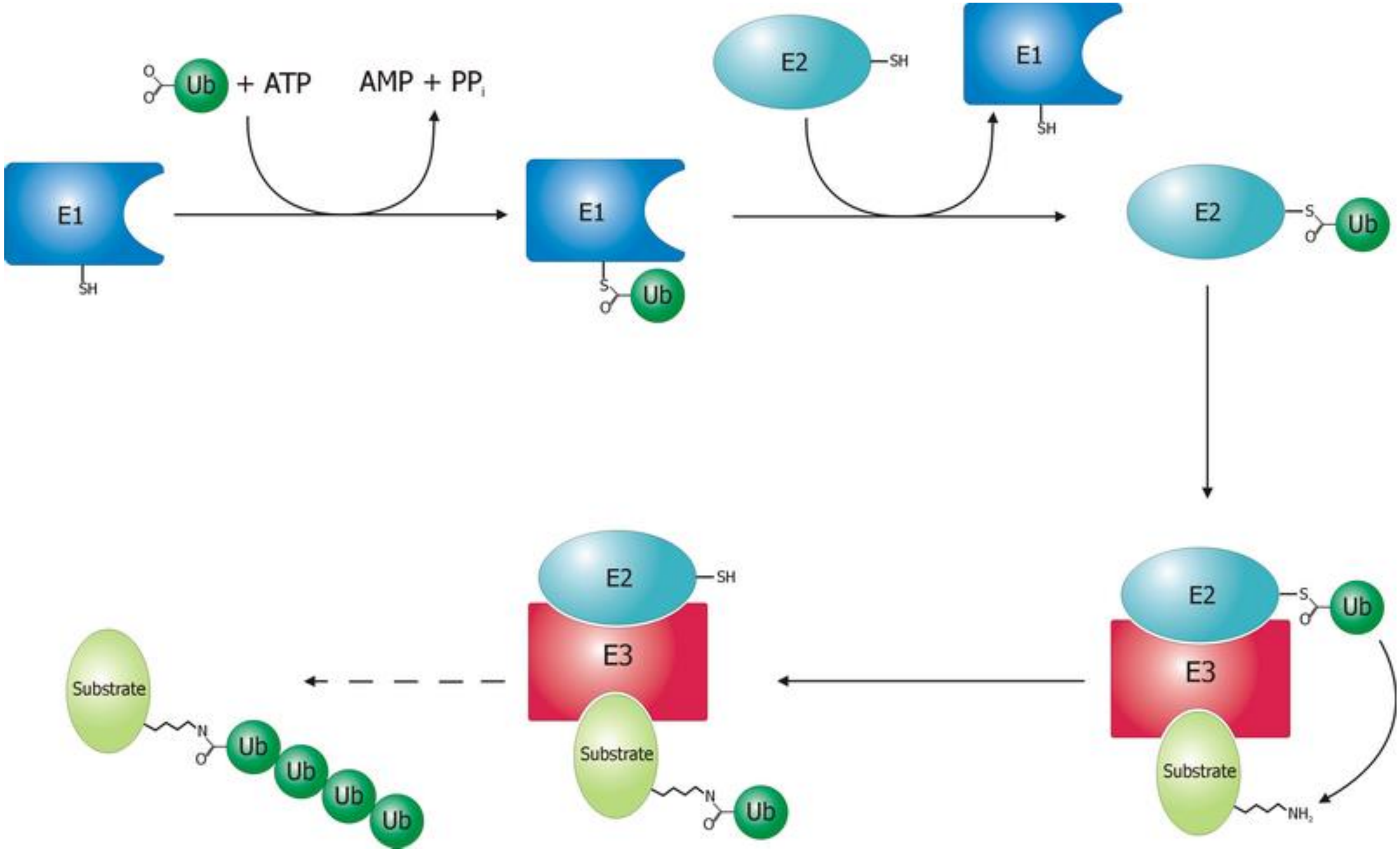
- Usually when searching these databases, you have either a region of DNA or a protein (or protein function) of interest.
- For this lab you'll be using a gene from Arabidopsis thaliana, a small flowering plant that is like the fruit fly of the plant world as it has a comparatively rapid life cycle and requires little space to grow.
- The protein product of this gene is recorded under accession number ***NP_001318308***, and it is an E3 ligase, involved in ubiquitination of proteins, which is a signal for their degradation.



https://vi.wikipedia.org/wiki/Arabidopsis_thaliana

Ubiquitin hóa

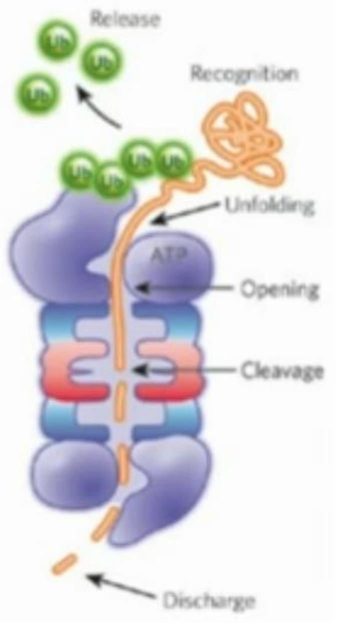
- Chỉ sự biến đổi sau dịch mã của một protein bằng cách gắn cộng hóa trị một hoặc nhiều đơn phân ubiquitin vào protein này.
- Đánh dấu các protein cho quá trình tiêu hủy. Quá trình tiêu hủy protein được thực hiện bởi bộ máy phân hủy protein (proteasome).
- Enzyme E1 khởi động quá trình này.
- Quá trình ubiquitin hóa cũng sẽ kiểm soát độ bền, chức năng và sự định vị nội bào của nhiều loại protein.



https://en.wikipedia.org/wiki/Ubiquitin_ligase

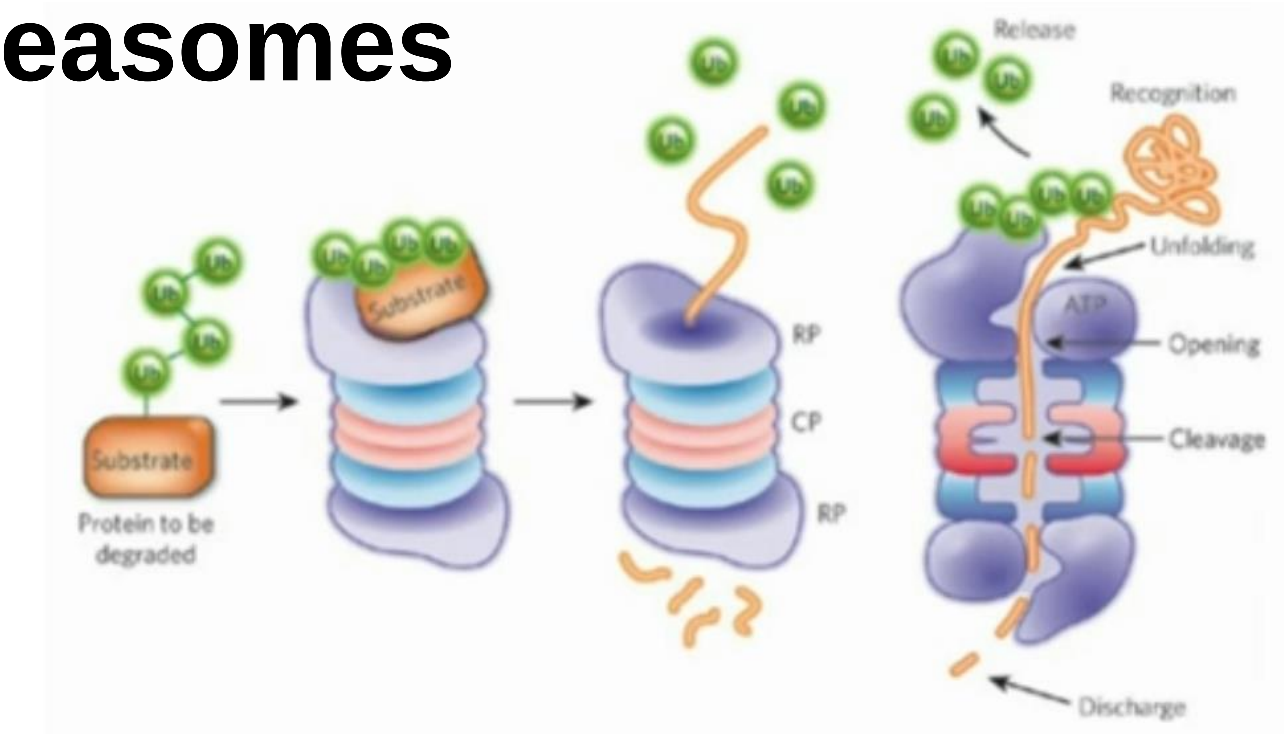
Proteasomes

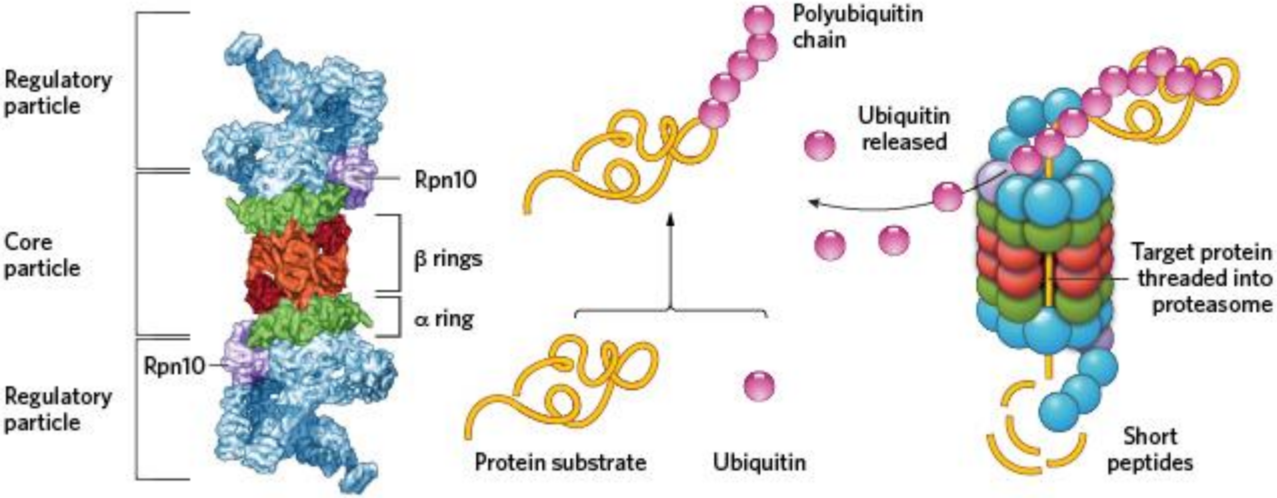
- **Proteasomes** là phức hợp protein bên trong tất cả các sinh vật nhân chuẩn và vi khuẩn cổ, và trong một số vi khuẩn.
- Ở eukaryote, chúng nằm trong nhân và tế bào chất[1]. Chức năng chính của các proteasome là tiêu hủy các protein không cần thiết hoặc bị hư hỏng bởi sự phân giải protein, một phản ứng hóa học phá vỡ liên kết peptit.
- Các enzym thực hiện phản ứng như vậy được gọi là protease.
- Proteasomes là một phần của một cơ chế chính mà nhờ đó các tế bào điều chỉnh nồng độ của protein đặc biệt và phân hủy protein cuộn sai.
- Quá trình này phá hủy tạo ra các peptide dài khoảng 7-8 amino acid, mà sau đó có thể được tiếp tục bị phá hủy tthành các amino acid và được sử dụng trong tổng hợp protein mới[2].



https://vi.wikipedia.org/wiki/Proteasome

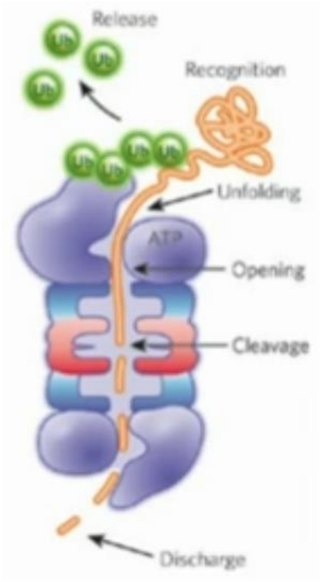
Proteasomes

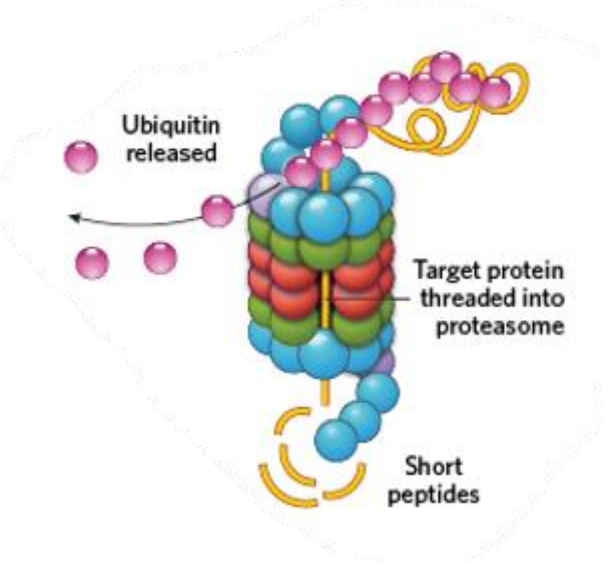




Proteasomes

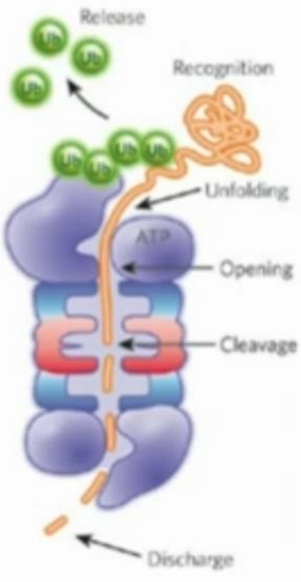
- Protein được đánh dấu để làm thoái biến bằng một loại protein nhỏ có tên là ubiquitin.
- Phản ứng đánh dấu được xúc tác bởi các enzyme gọi là ubiquitin ligases.
- Một khi một protein được đánh dấu bằng một phân tử ubiquitin duy nhất, đây là một tín hiệu ligases khác để gắn các phân tử ubiquitin thêm.
- Kết quả là một chuỗi polyubiquitin bị ràng buộc bởi proteasome, cho phép nó làm suy biến protein gắn thẻ.

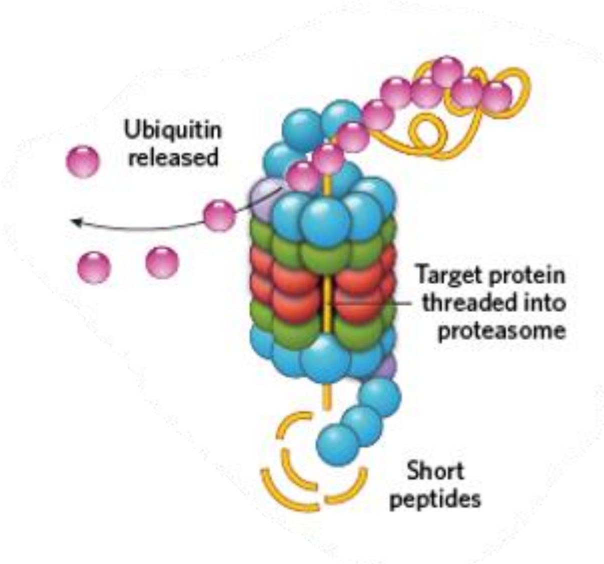




Proteasomes

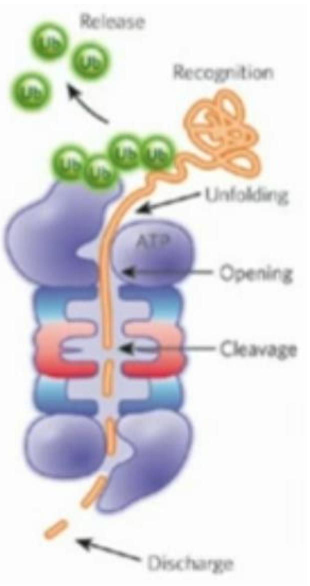
- Về cấu trúc, proteasome là một phức hợp hình trụ có chứa một "lõi" của bốn vòng tròn xếp chồng lên nhau xung quanh một lỗ trung tâm.
- Mỗi vòng gồm bảy protein riêng lẻ.
- Hai vòng trong gồm bảy tiểu đơn vị β có chứa 3-7 các địa điểm hoạt động của protease.
- Những địa điểm này được đặt trên bề mặt bên trong của những vòng nhẫn, để các protein mục tiêu phải đi vào lỗ trung tâm trước khi nó bị suy biến.
- Hai vòng ngoài từng có 7 tiểu đơn vị α có chức năng là để duy trì một "cửa" mà thông qua đó các protein vào ống.
- Các tiểu đơn vị α được kiểm soát bằng cách liên kết với "mũ" cấu trúc hoặc các hạt quản lý nhận ra thẻ polyubiquitin gắn liền với chất nền protein và bắt đầu quá trình suy biến.

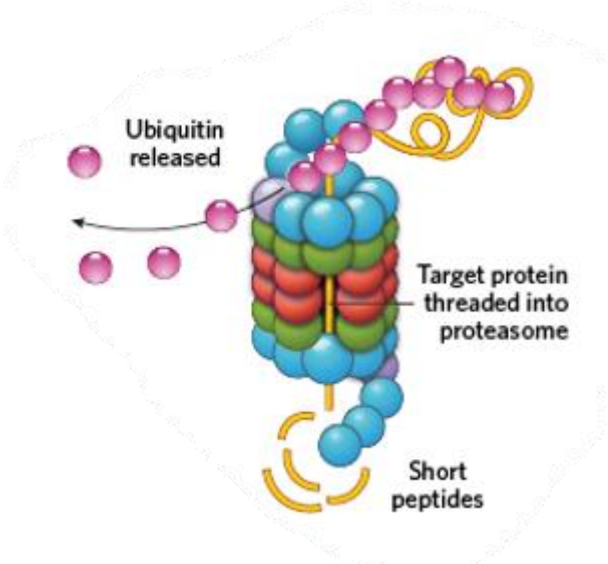


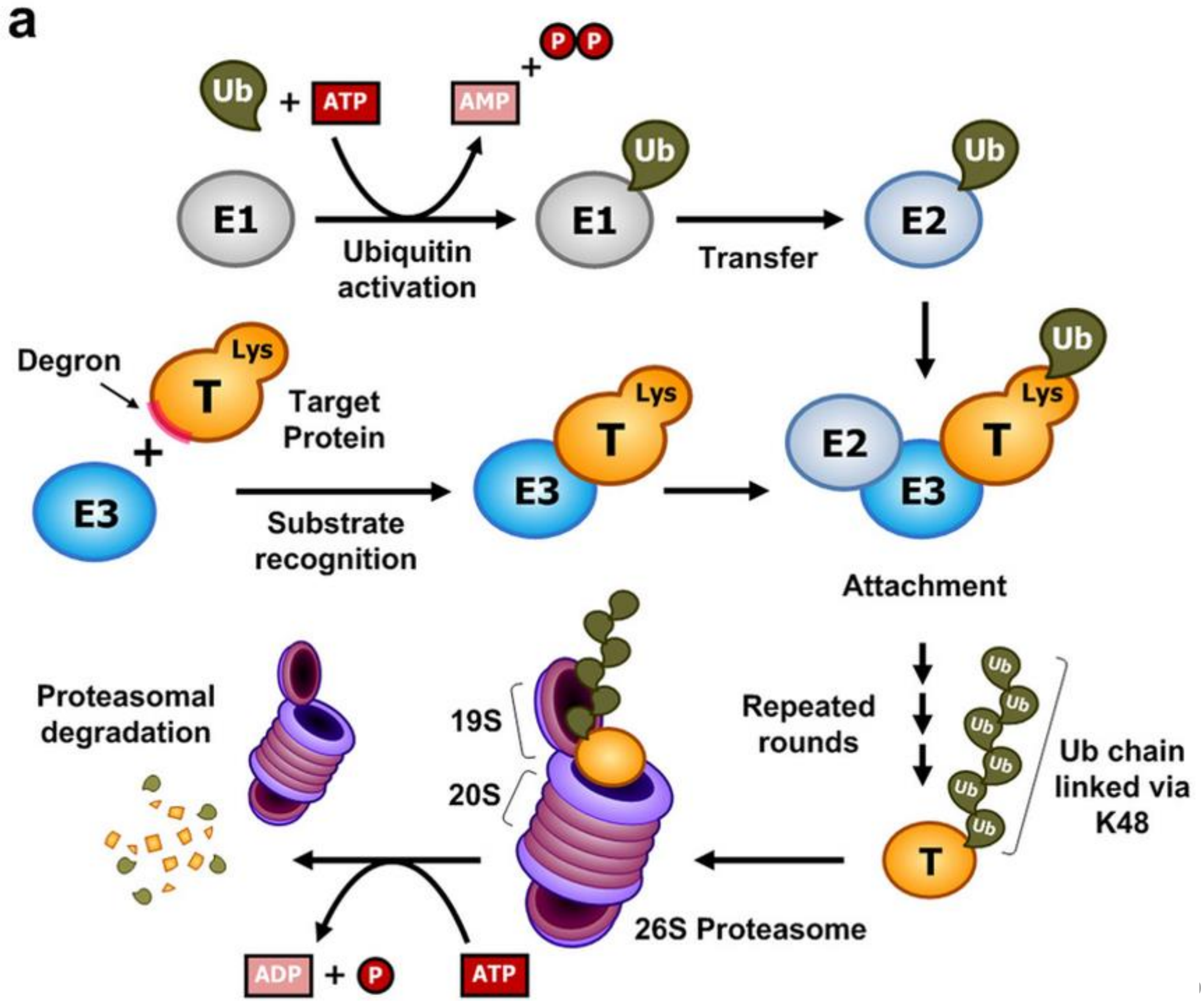


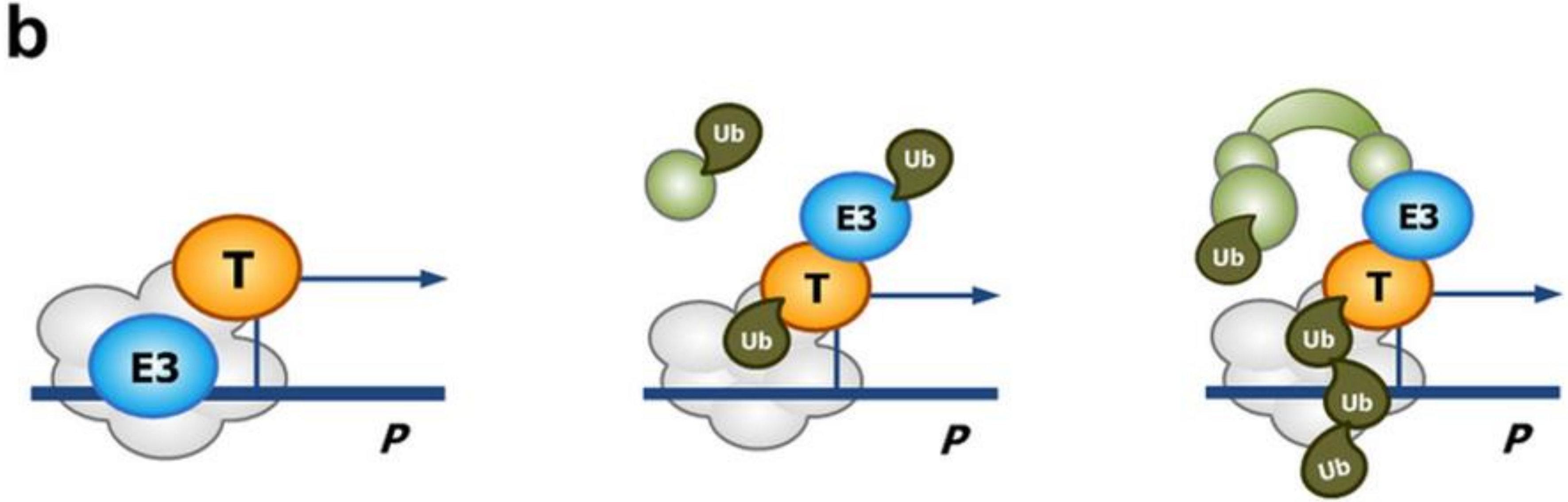
Proteasomes

- Toàn bộ hệ thống của ubiquitination và suy biến proteasome được biết đến với tên hệ thống ubiquitin-proteasome.
- Các con đường biến proteasome là điều cần thiết cho nhiều quá trình tế bào, bao gồm cả chu kỳ tế bào, quy định biểu hiện gen, và phản ứng với stress oxy hóa.
- Tầm quan trọng của suy thoái thủy phân protein bên trong tế bào và vai trò của ubiquitin trong con đường thủy phân protein đã được thừa nhận khi trao giải Nobel Hóa học năm 2004 cho Aaron Ciechanover, Avram Hershko và Irwin Rose[3].






a
Ub
+
ATP
AMP
+
P P
E1
Ubiquitin activation
E1
Ub
Transfer
E2
Ub
Degron
T
Lys
Target Protein
+
E3
Substrate recognition
E3
T
Lys
E2
E3
T
Lys
Ub
Attachment
Repeated rounds
Ub
Ub
Ub
Ub
Ub
Ub
T
Ub chain linked via K48
Proteasomal degradation
19S
20S
26S Proteasome
ADP
+
P
ATP



Cofactor

Mono-ubiquitination

Poly-ubiquitination

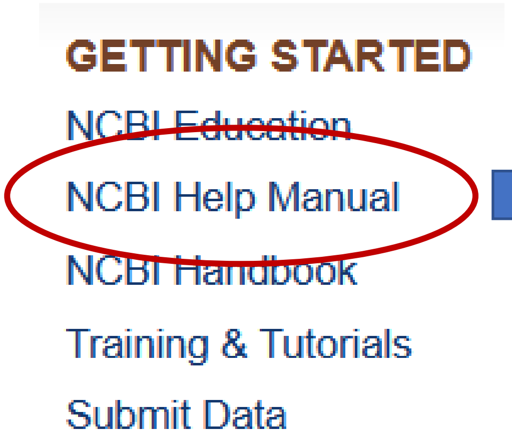
Bước 4. Tìm kiếm Protein

- Thực hiện các thao tác tìm kiếm sau. ***Chụp lại màn hình kết quả tìm kiếm và nhận xét.***
- ***gene keywords***: e.g. ubiquitin-protein ligase
- ***gene keyword AND organism***: e.g. ubiquitin-protein ligase AND Arabidopsis thaliana
- ***gene keyword [PROT] AND organism [ORGN]***: e.g. ubiquitin-protein ligase [PROT] AND Arabidopsis thaliana [ORGN]
- ***accession or GI number***: e.g. NP_001318308

Thực hành: sử dụng NCBI Help

1. At the bottom left of the NCBI homepage find the "NCBI Help Manual" link. Click on it.

Then access the "Entrez Help" section.



GETTING STARTED

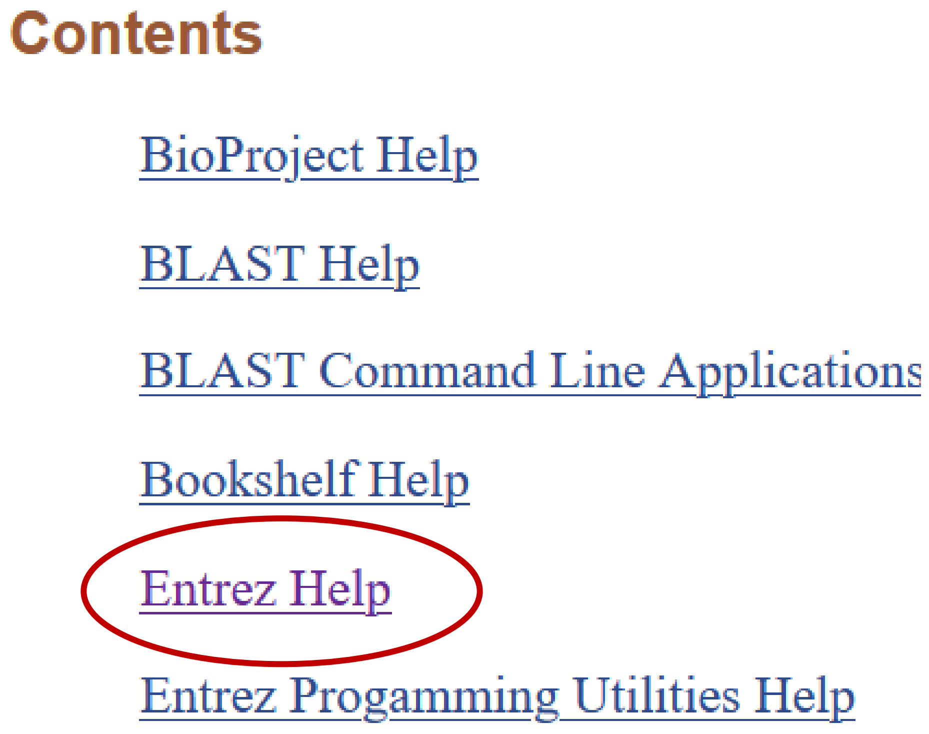
NCBI Education

NCBI Help Manual

NCBI Handbook

Training & Tutorials

Submit Data



Contents

BioProject Help

BLAST Help

BLAST Command Line Applications

Bookshelf Help

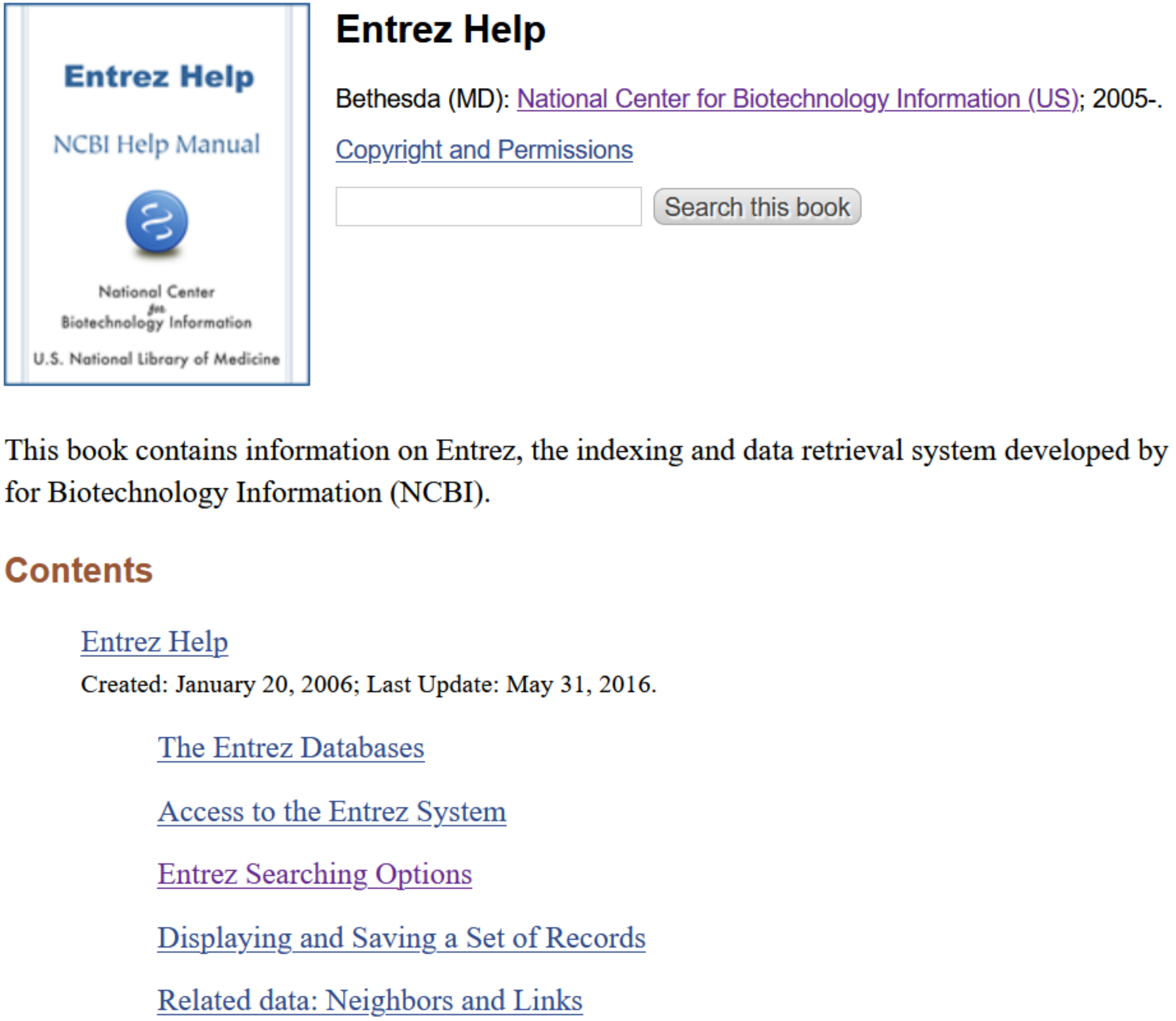
Entrez Help

Entrez Progamming Utilities Help

https://www.ncbi.nlm.nih.gov/books/NBK3831/

2. You are now in Entrez Help.

- The Entrez collection of databases is queried when you use the Search NCBI interface.
- Note the contents that explain everything from search options to saving sets of records.


Entrez Help
NCBI Help Manual
National Center for Biotechnology Information
U.S. National Library of Medicine
Search this book

Entrez Help

Bethesda (MD): National Center for Biotechnology Information (US); 2005-.

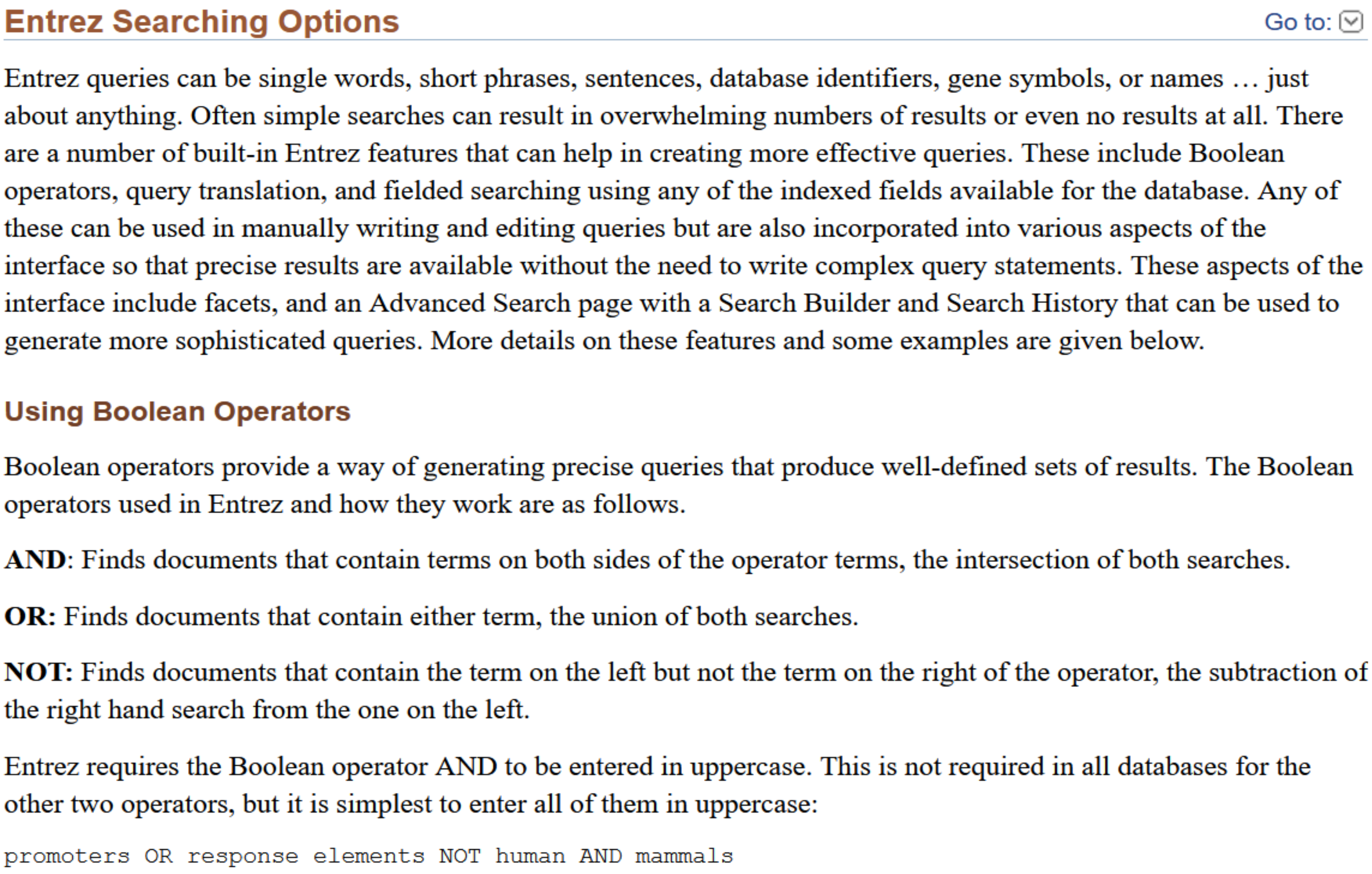
Copyright and Permissions

This book contains information on Entrez, the indexing and data retrieval system developed by for Biotechnology Information (NCBI).

Contents

https://www.ncbi.nlm.nih.gov/books/NBK3836/

3. Notice that under the section Entrez Searching Options some other appropriate qualifiers are given, as illustrated on the previous section.



Entrez Searching Options

Entrez queries can be single words, short phrases, sentences, database identifiers, gene symbols, or names … just about anything. Often simple searches can result in overwhelming numbers of results or even no results at all. There are a number of built-in Entrez features that can help in creating more effective queries. These include Boolean operators, query translation, and fielded searching using any of the indexed fields available for the database. Any of these can be used in manually writing and editing queries but are also incorporated into various aspects of the interface so that precise results are available without the need to write complex query statements. These aspects of the interface include facets, and an Advanced Search page with a Search Builder and Search History that can be used to generate more sophisticated queries. More details on these features and some examples are given below.

Using Boolean Operators

Boolean operators provide a way of generating precise queries that produce well-defined sets of results. The Boolean operators used in Entrez and how they work are as follows.

AND: Finds documents that contain terms on both sides of the operator terms, the intersection of both searches.

OR: Finds documents that contain either term, the union of both searches.

NOT: Finds documents that contain the term on the left but not the term on the right of the operator, the subtraction of the right hand search from the one on the left.

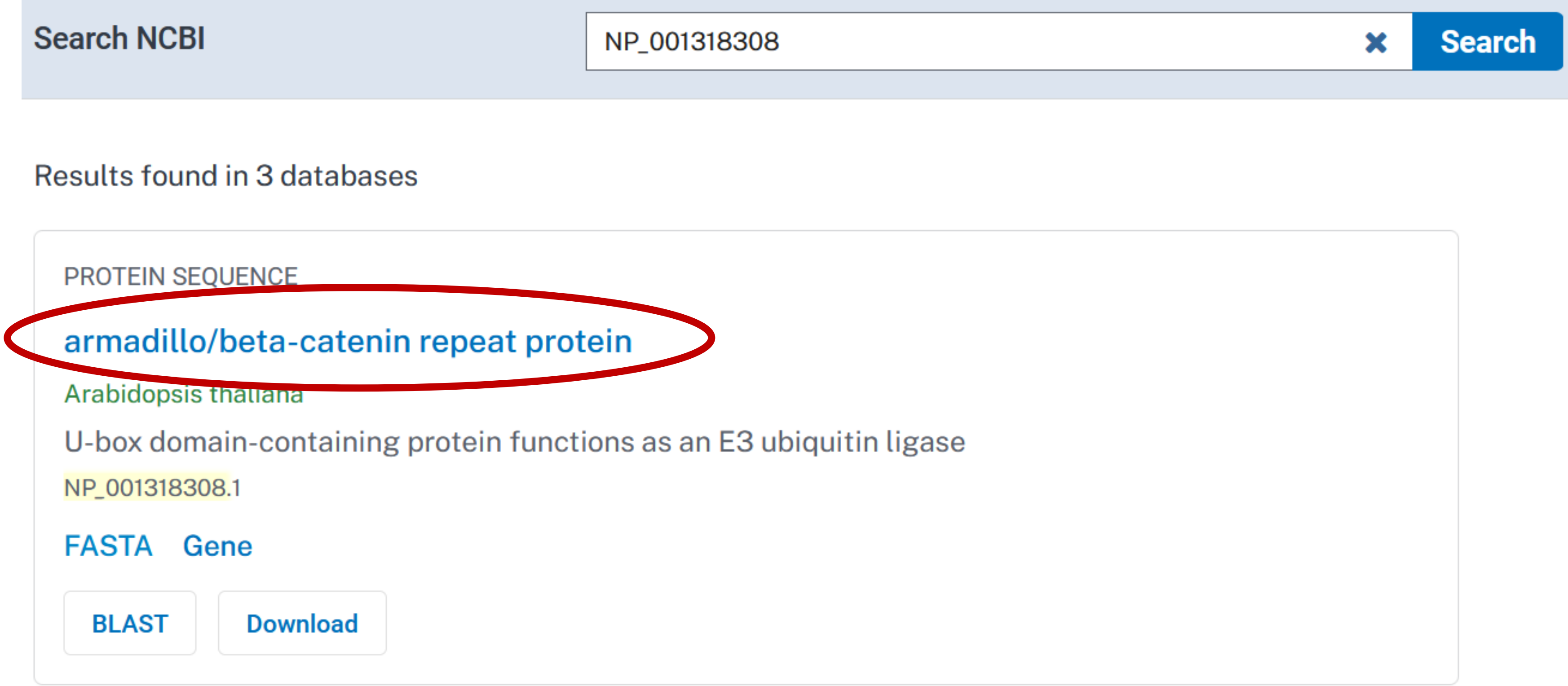
Entrez requires the Boolean operator AND to be entered in uppercase. This is not required in all databases for the other two operators, but it is simplest to enter all of them in uppercase:

```
promoters OR response elements NOT human AND mammals
```

https://www.ncbi.nlm.nih.gov/books/NBK3837/#EntrezHelp.Entrez_Searching_Options

Bước 5.

- Search for our accession number of interest (e.g. NP_001318308 from above) through the Search NCBI portal page.



Search NCBI

NP_001318308

Results found in 3 databases

PROTEIN SEQUENCE

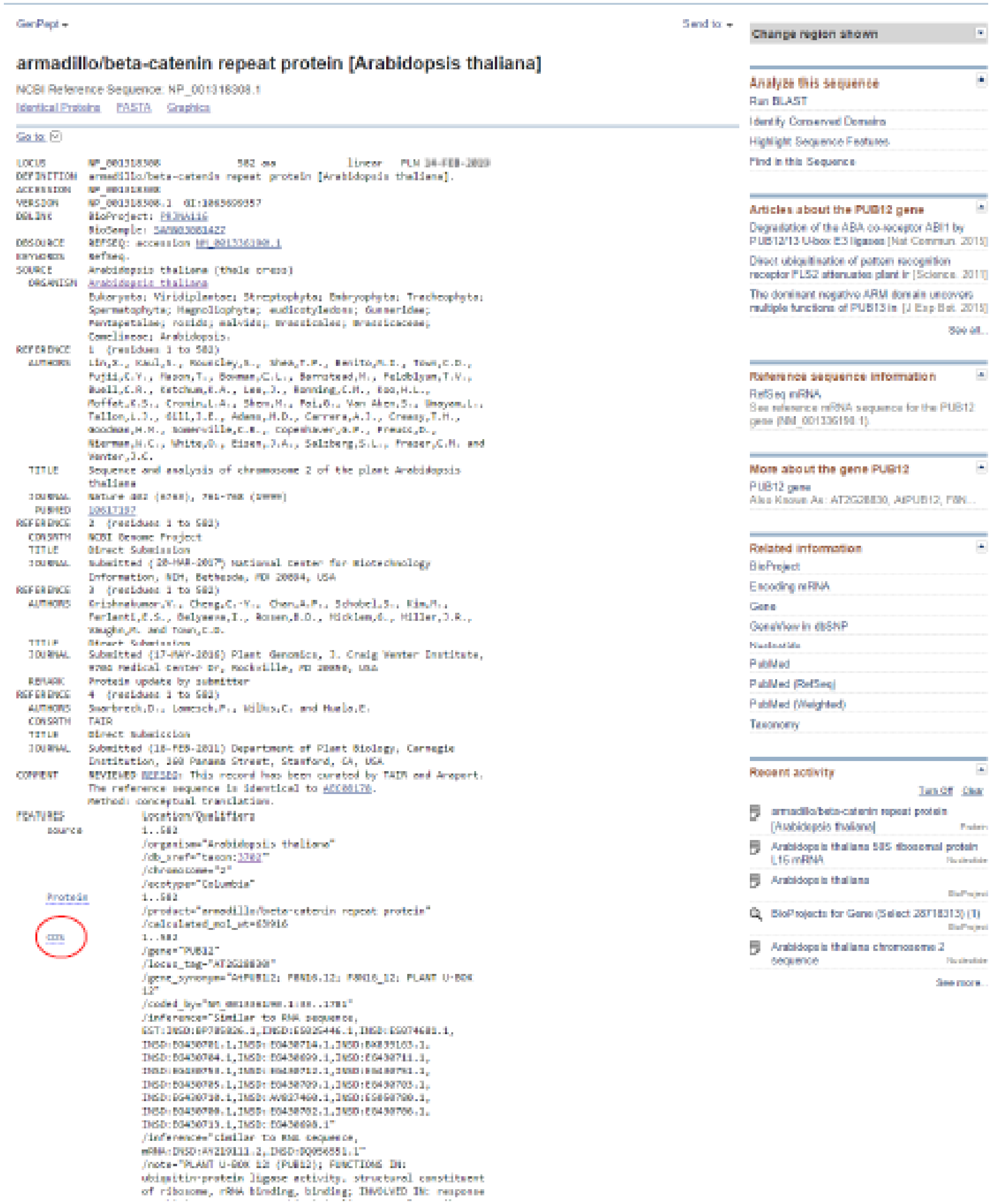
armadillo/beta-catenin repeat protein

Arabidopsis thaliana

U-box domain-containing protein functions as an E3 ubiquitin ligase

NP_001318308.1

https://www.ncbi.nlm.nih.gov/search/all/?term=NP_001318308%20



GenPept ▾ Send to: ▾

armadillo/beta-catenin repeat protein [Arabidopsis thaliana]

NCBI Reference Sequence: NP_001318308.1

Identical Proteins FASTA Graphics

Go to: ☑

```
LOCUS       NP_001318308             582 aa            linear   PLN [illegible]
DEFINITION  armadillo/beta-catenin repeat protein [Arabidopsis thaliana].
ACCESSION   NP_001318308
VERSION     NP_001318308.1  GI:[illegible]
DBLINK      BioProject: PRJNA116
            BioSample: SAMN03081427
DBSOURCE    REFSEQ: accession NM_001336198.1
KEYWORDS    RefSeq.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae;
            Pentapetalae; rosids; malvids; Brassicales; Brassicaceae;
            Camelineae; Arabidopsis.
REFERENCE   1  (residues 1 to 582)
  AUTHORS   Lin,X., Kaul,S., Rounsley,S., Shea,T.P., Benito,M.I., Town,C.D.,
            Fujii,C.Y., Mason,T., Bowman,C.L., Barnstead,M., Feldblyum,T.V.,
            Buell,C.R., Ketchum,K.A., Lee,J., Ronning,C.M., Koo,H.L.,
            Moffat,K.S., Cronin,L.A., Shen,M., Pai,G., Van Aken,S., Umayam,L.,
            Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H.,
            Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D.,
            Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and
            Venter,J.C.
  TITLE     Sequence and analysis of chromosome 2 of the plant Arabidopsis
            thaliana
  JOURNAL   Nature 402 (6763), 761-768 (1999)
   PUBMED   10617197
REFERENCE   2  (residues 1 to 582)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-2017) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 582)
  AUTHORS   Krishnakumar,V., Cheng,C.-Y., Chan,A.P., Schobel,S., Kim,M.,
            Ferlanti,E.S., Belyaeva,I., Rosen,B.D., Micklem,G., Miller,J.R.,
            Vaughn,M. and Town,C.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-2016) Plant Genomics, J. Craig Venter Institute,
            9704 Medical Center Dr, Rockville, MD 20850, USA
  REMARK    Protein update by submitter
REFERENCE   4  (residues 1 to 582)
  AUTHORS   Swarbreck,D., Lamesch,P., Wilks,C. and Huala,E.
  CONSRTM   TAIR
  TITLE     Direct Submission
  JOURNAL   Submitted (18-FEB-2011) Department of Plant Biology, Carnegie
            Institution, 260 Panama Street, Stanford, CA, USA
COMMENT     REVIEWED REFSEQ: This record has been curated by TAIR and Araport.
            The reference sequence is identical to AEC08178.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..582
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /ecotype="Columbia"
     Protein         1..582
                     /product="armadillo/beta-catenin repeat protein"
                     /calculated_mol_wt=63916
     CDS             1..582
                     /gene="PUB12"
                     /locus_tag="AT2G28830"
                     /gene_synonym="AtPUB12; F8N16.12; F8N16_12; PLANT U-BOX
                     12"
                     /coded_by="NM_001336198.1:[illegible]..1781"
                     /inference="Similar to RNA sequence,
                     EST:INSD:BP785826.1,INSD:ES025446.1,INSD:ES074681.1,
                     INSD:[illegible]
                     INSD:[illegible]
                     INSD:[illegible]
                     INSD:[illegible]
                     INSD:[illegible]
                     INSD:[illegible]
                     INSD:[illegible]"
                     /inference="similar to RNA sequence,
                     mRNA:INSD:AY219111.2,INSD:BQ056551.1"
                     /note="PLANT U-BOX 12 (PUB12); FUNCTIONS IN:
                     ubiquitin-protein ligase activity, structural constituent
                     of ribosome, rRNA binding, binding; INVOLVED IN: response
```

Change region shown

Analyze this sequence

- Run BLAST
- Identify Conserved Domains
- Highlight Sequence Features
- Find in this Sequence

Articles about the PUB12 gene

- Degradation of the ABA co-receptor ABI1 by PUB12/13 U-box E3 ligases [Nat Commun. 2015]
- Direct ubiquitination of pattern recognition receptor FLS2 attenuates plant in [Science. 2011]
- The dominant negative ARM domain uncovers multiple functions of PUB13 in [J Exp Bot. 2015]

See all...

Reference sequence information

RefSeq mRNA
See reference mRNA sequence for the PUB12 gene (NM_001336198.1).

More about the gene PUB12

PUB12 gene
Also Known As: AT2G28830, AtPUB12, F8N...

Related information

- BioProject
- Encoding mRNA
- Gene
- GeneView in dbSNP
- Nucleotide
- PubMed
- PubMed (RefSeq)
- PubMed (Weighted)
- Taxonomy

Recent activity

Turn Off Clear

- armadillo/beta-catenin repeat protein [Arabidopsis thaliana] Protein
- Arabidopsis thaliana 60S ribosomal protein L15 mRNA Nucleotide
- Arabidopsis thaliana BioProject
- BioProjects for Gene (Select 28718313) (1) BioProject
- Arabidopsis thaliana chromosome 2 sequence Nucleotide

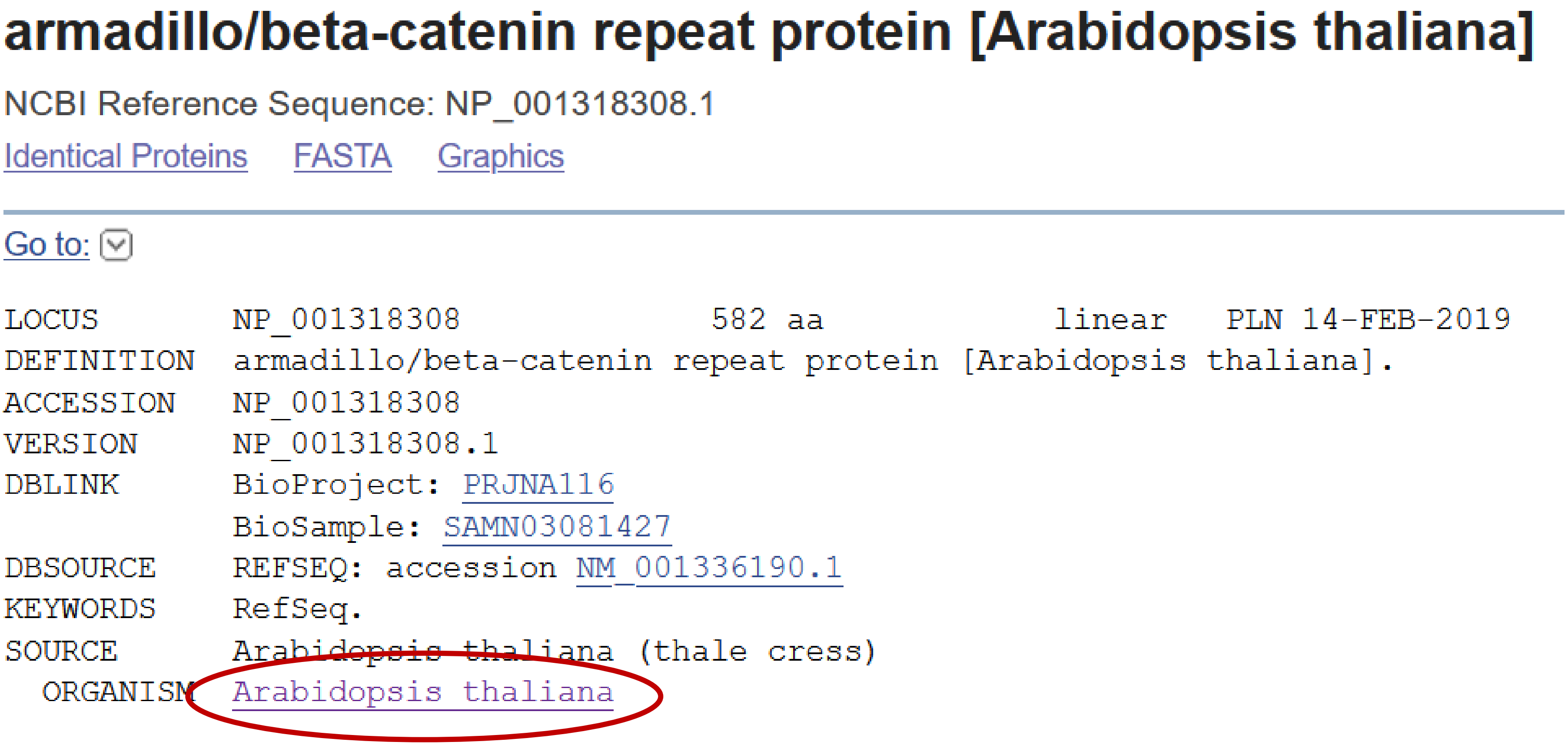
See more...

https://www.ncbi.nlm.nih.gov/protein/NP_001318308.1/

```
                     (InterPro:IPR016180), Armadillo (InterPro:IPR000225),
```

Bước 6.

- Click on the SOURCE ORGANISM hyperlink.



armadillo/beta-catenin repeat protein [Arabidopsis thaliana]

NCBI Reference Sequence: NP_001318308.1

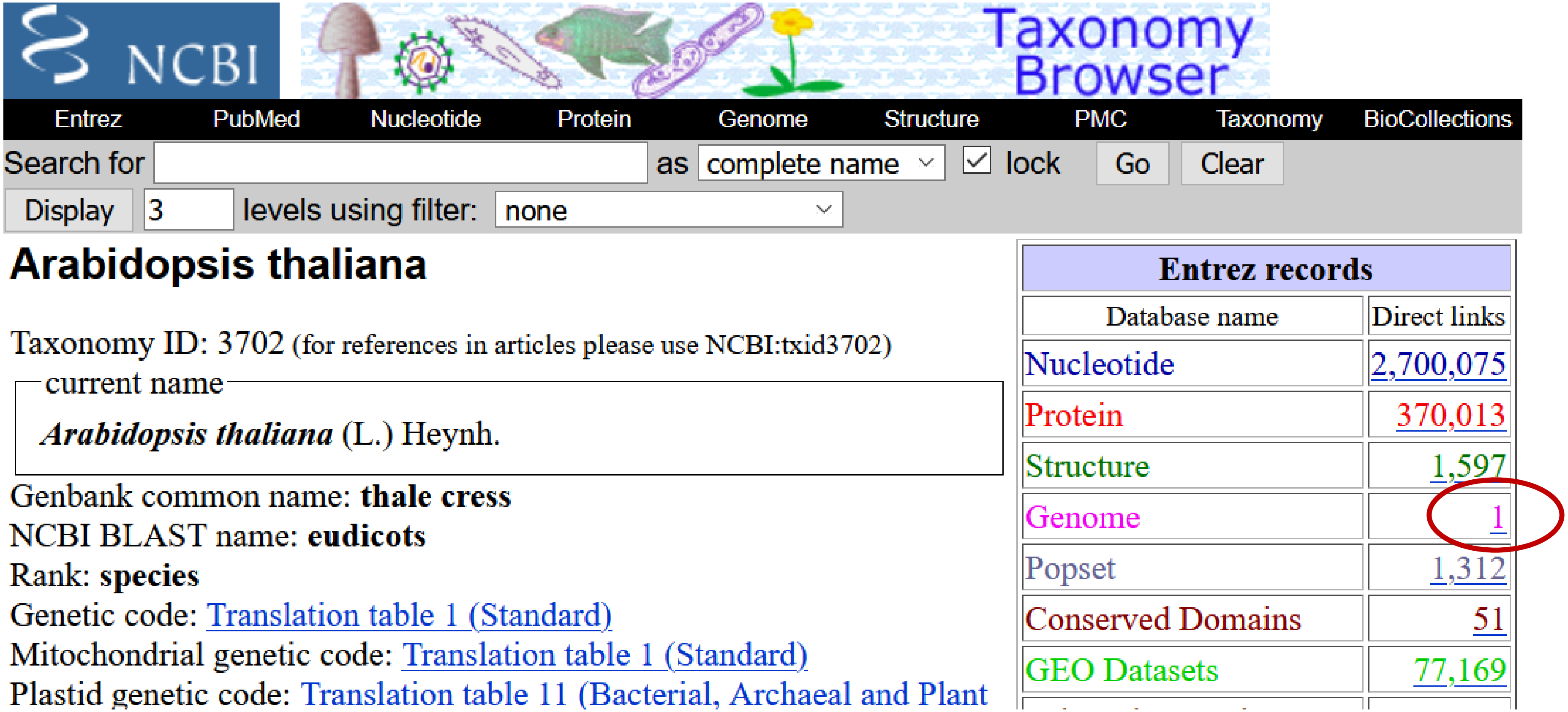
Identical Proteins FASTA Graphics

Go to:

```
LOCUS       NP_001318308            582 aa            linear   PLN 14-FEB-2019
DEFINITION  armadillo/beta-catenin repeat protein [Arabidopsis thaliana].
ACCESSION   NP_001318308
VERSION     NP_001318308.1
DBLINK      BioProject: PRJNA116
            BioSample: SAMN03081427
DBSOURCE    REFSEQ: accession NM_001336190.1
KEYWORDS    RefSeq.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
```

https://www.ncbi.nlm.nih.gov/protein/NP_001318308.1/

Ấn vào link Genome



Arabidopsis thaliana

Taxonomy ID: 3702 (for references in articles please use NCBI:txid3702)

current name

Arabidopsis thaliana (L.) Heynh.

Genbank common name: **thale cress**
NCBI BLAST name: **eudicots**
Rank: **species**
Genetic code: Translation table 1 (Standard)
Mitochondrial genetic code: Translation table 1 (Standard)
Plastid genetic code: Translation table 11 (Bacterial, Archaeal and Plant

Entrez records	
Database name	Direct links
Nucleotide	2,700,075
Protein	370,013
Structure	1,597
Genome	1
Popset	1,312
Conserved Domains	51
GEO Datasets	77,169

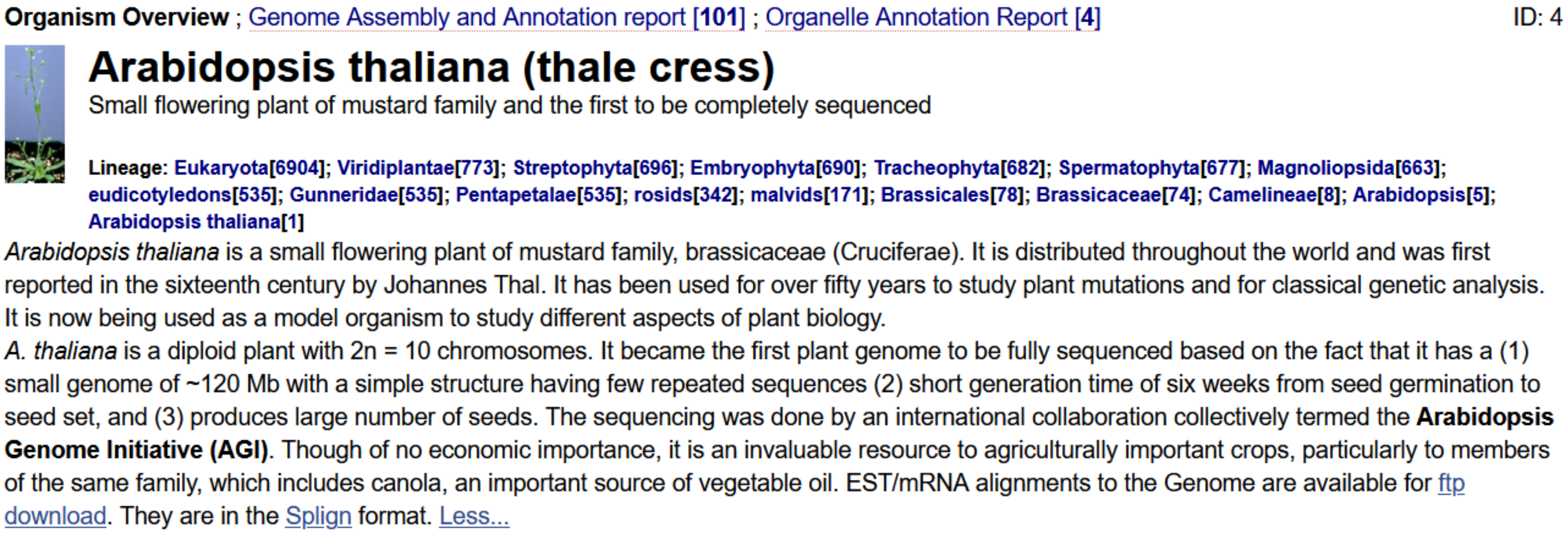
https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=3702

Trả lời các câu hỏi sau:

6a. What is the taxonomic lineage of your organism?

6b. Has the genome of this organism been sequenced, i.e. is there a Genome Project?

6c. If so, can you find the accession for the full sequence or one of the chromosomes?



Organism Overview ; Genome Assembly and Annotation report [**101**] ; Organelle Annotation Report [**4**] ID: 4

Arabidopsis thaliana (thale cress)

Small flowering plant of mustard family and the first to be completely sequenced

Lineage: Eukaryota[6904]; Viridiplantae[773]; Streptophyta[696]; Embryophyta[690]; Tracheophyta[682]; Spermatophyta[677]; Magnoliopsida[663]; eudicotyledons[535]; Gunneridae[535]; Pentapetalae[535]; rosids[342]; malvids[171]; Brassicales[78]; Brassicaceae[74]; Camelineae[8]; Arabidopsis[5]; Arabidopsis thaliana[1]

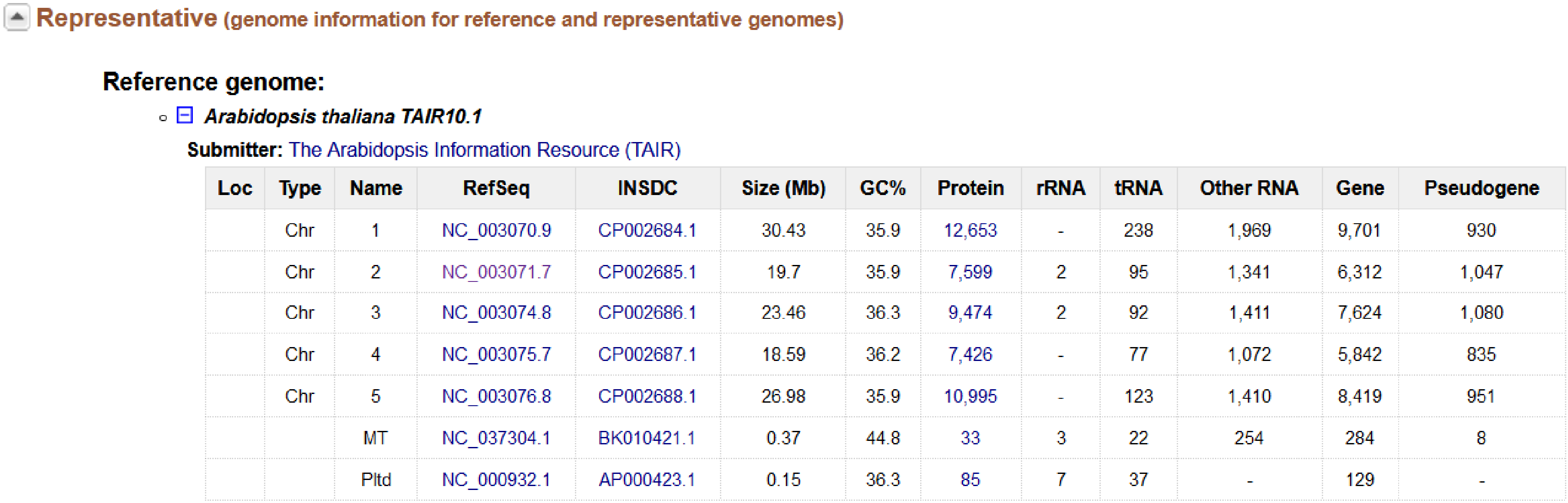
Arabidopsis thaliana is a small flowering plant of mustard family, brassicaceae (Cruciferae). It is distributed throughout the world and was first reported in the sixteenth century by Johannes Thal. It has been used for over fifty years to study plant mutations and for classical genetic analysis. It is now being used as a model organism to study different aspects of plant biology.

A. thaliana is a diploid plant with 2n = 10 chromosomes. It became the first plant genome to be fully sequenced based on the fact that it has a (1) small genome of ~120 Mb with a simple structure having few repeated sequences (2) short generation time of six weeks from seed germination to seed set, and (3) produces large number of seeds. The sequencing was done by an international collaboration collectively termed the **Arabidopsis Genome Initiative (AGI)**. Though of no economic importance, it is an invaluable resource to agriculturally important crops, particularly to members of the same family, which includes canola, an important source of vegetable oil. EST/mRNA alignments to the Genome are available for ftp download. They are in the Splign format. Less...

- *Arabidopsis thaliana* is a small flowering plant of mustard family, brassicaceae (Cruciferae).
- It is distributed throughout the world and was first reported in the sixteenth century by Johannes Thal.
- It has been used for over fifty years to study plant mutations and for classical genetic analysis.
- It is now being used as a model organism to study different aspects of plant biology.

https://www.ncbi.nlm.nih.gov/genome/?term=txid3702[Organism:noexp]

- *A. thaliana* is a diploid plant with 2n = 10 chromosomes.
- It became the first plant genome to be fully sequenced based on the fact that it has a (1) small genome of ~120 Mb with a simple structure having few repeated sequences (2) short generation time of six weeks from seed germination to seed set, and (3) produces large number of seeds.
- ***Answer 6b.*** The sequencing was done by an international collaboration collectively termed the **Arabidopsis Genome Initiative (AGI)**.
- Though of no economic importance, it is an invaluable resource to agriculturally important crops, particularly to members of the same family, which includes canola, an important source of vegetable oil. EST/mRNA alignments to the Genome are available for ftp download.
- They are in the Splign format.



Representative (genome information for reference and representative genomes)

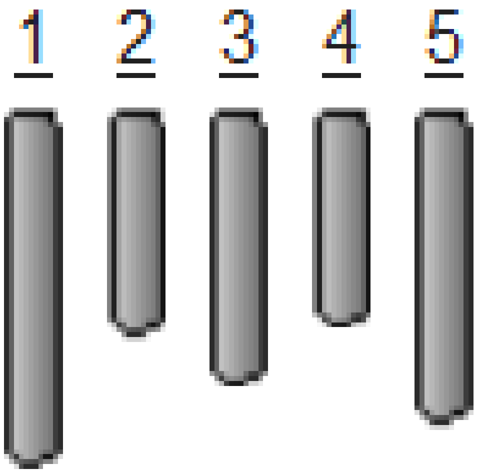
Reference genome:

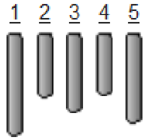
- ⊟ ***Arabidopsis thaliana TAIR10.1***

Submitter: The Arabidopsis Information Resource (TAIR)

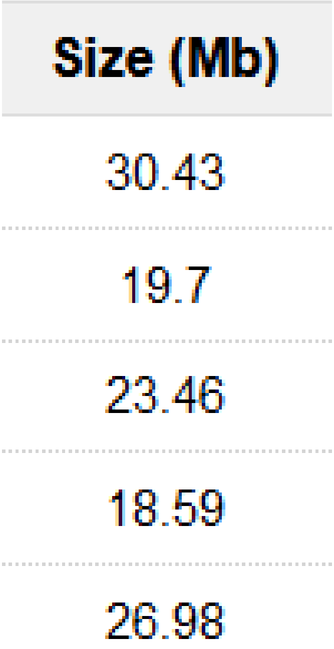
Loc	Type	Name	RefSeq	INSDC	Size (Mb)	GC%	Protein	rRNA	tRNA	Other RNA	Gene	Pseudogene
	Chr	1	NC_003070.9	CP002684.1	30.43	35.9	12,653	-	238	1,969	9,701	930
	Chr	2	NC_003071.7	CP002685.1	19.7	35.9	7,599	2	95	1,341	6,312	1,047
	Chr	3	NC_003074.8	CP002686.1	23.46	36.3	9,474	2	92	1,411	7,624	1,080
	Chr	4	NC_003075.7	CP002687.1	18.59	36.2	7,426	-	77	1,072	5,842	835
	Chr	5	NC_003076.8	CP002688.1	26.98	35.9	10,995	-	123	1,410	8,419	951
		MT	NC_037304.1	BK010421.1	0.37	44.8	33	3	22	254	284	8
		Pltd	NC_000932.1	AP000423.1	0.15	36.3	85	7	37	-	129	-

Chromosomes





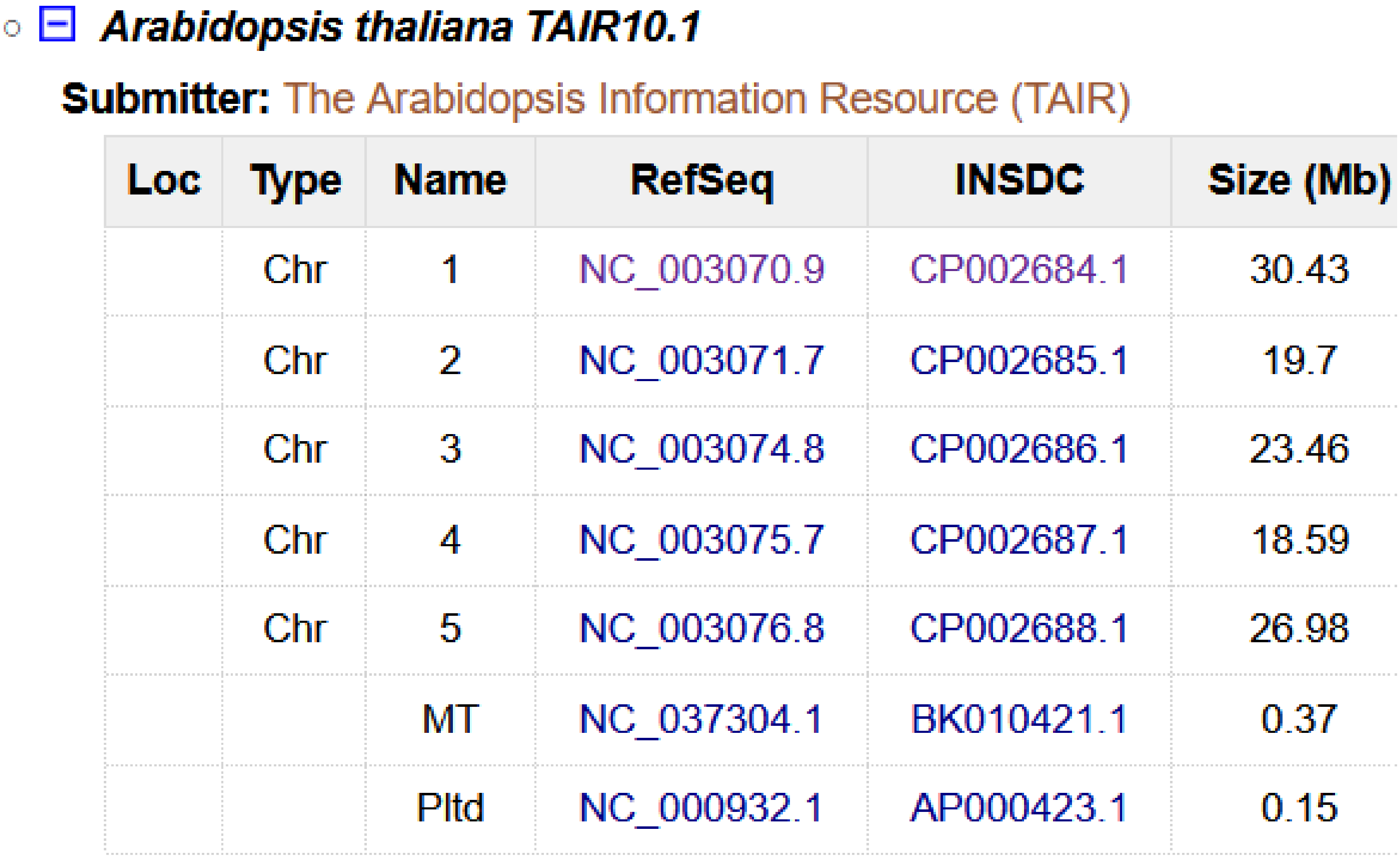
Click on chromosome name to open Genome Data Viewer



Size (Mb)
30.43
19.7
23.46
18.59
26.98

https://www.ncbi.nlm.nih.gov/genome/?term=txid3702[Organism:noexp]

The Arabidopsis Information Resource (TAIR)



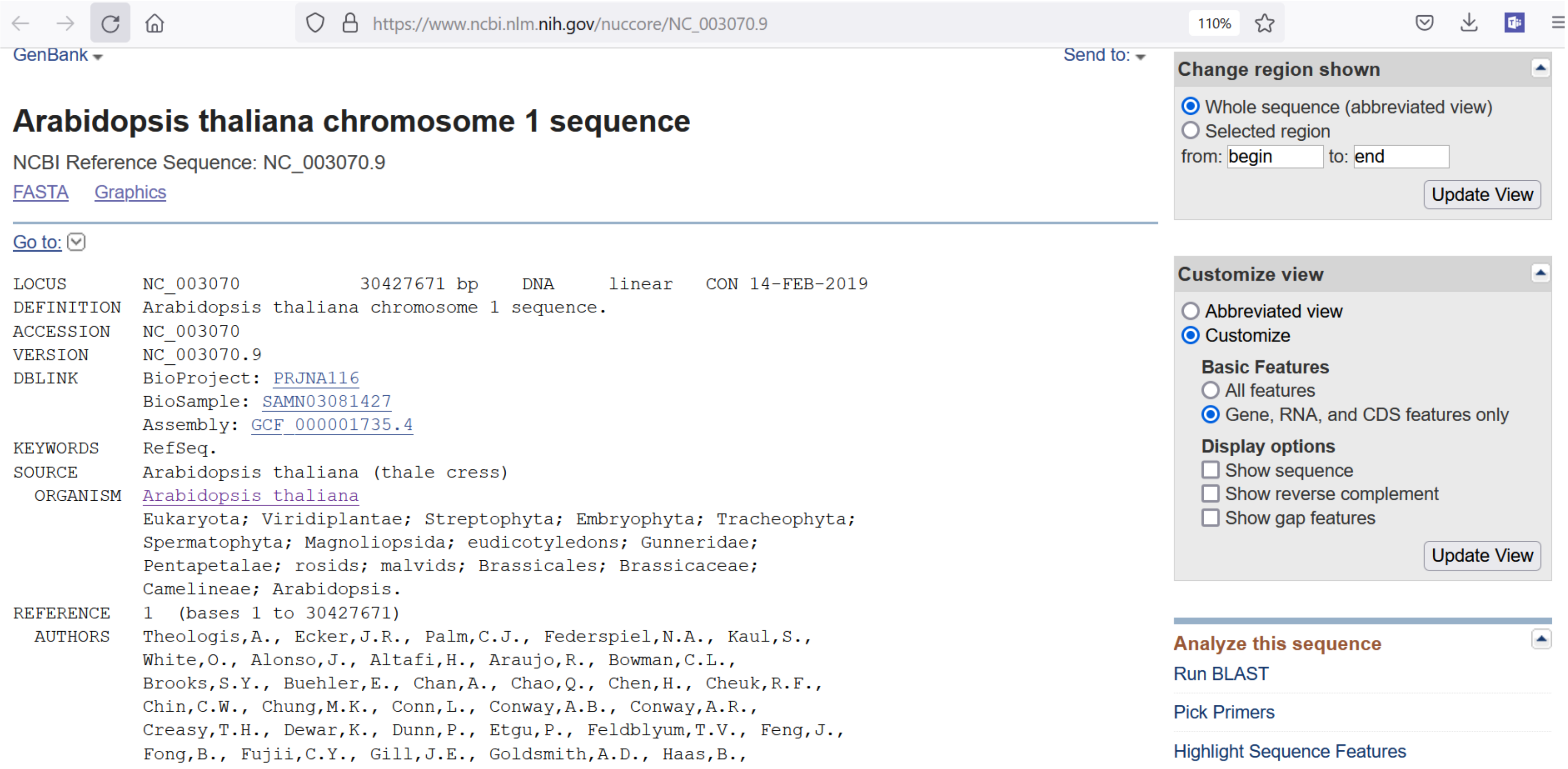
○ ⊟ ***Arabidopsis thaliana TAIR10.1***

Submitter: The Arabidopsis Information Resource (TAIR)

Loc	Type	Name	RefSeq	INSDC	Size (Mb)
	Chr	1	NC_003070.9	CP002684.1	30.43
	Chr	2	NC_003071.7	CP002685.1	19.7
	Chr	3	NC_003074.8	CP002686.1	23.46
	Chr	4	NC_003075.7	CP002687.1	18.59
	Chr	5	NC_003076.8	CP002688.1	26.98
		MT	NC_037304.1	BK010421.1	0.37
		Pltd	NC_000932.1	AP000423.1	0.15

Answer 6c

6c. If so, can you find the accession for the full sequence or one of the chromosomes?



https://www.ncbi.nlm.nih.gov/nuccore/NC_003070.9

Arabidopsis thaliana chromosome 1 sequence

NCBI Reference Sequence: NC_003070.9

```
LOCUS       NC_003070           30427671 bp    DNA     linear   CON 14-FEB-2019
DEFINITION  Arabidopsis thaliana chromosome 1 sequence.
ACCESSION   NC_003070
VERSION     NC_003070.9
DBLINK      BioProject: PRJNA116
            BioSample: SAMN03081427
            Assembly: GCF_000001735.4
KEYWORDS    RefSeq.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliopsida; eudicotyledons; Gunneridae;
            Pentapetalae; rosids; malvids; Brassicales; Brassicaceae;
            Camelineae; Arabidopsis.
REFERENCE   1  (bases 1 to 30427671)
  AUTHORS   Theologis,A., Ecker,J.R., Palm,C.J., Federspiel,N.A., Kaul,S.,
            White,O., Alonso,J., Altafi,H., Araujo,R., Bowman,C.L.,
            Brooks,S.Y., Buehler,E., Chan,A., Chao,Q., Chen,H., Cheuk,R.F.,
            Chin,C.W., Chung,M.K., Conn,L., Conway,A.B., Conway,A.R.,
            Creasy,T.H., Dewar,K., Dunn,P., Etgu,P., Feldblyum,T.V., Feng,J.,
            Fong,B., Fujii,C.Y., Gill,J.E., Goldsmith,A.D., Haas,B.,
```

Change region shown

● Whole sequence (abbreviated view)
○ Selected region
from: begin to: end

Customize view

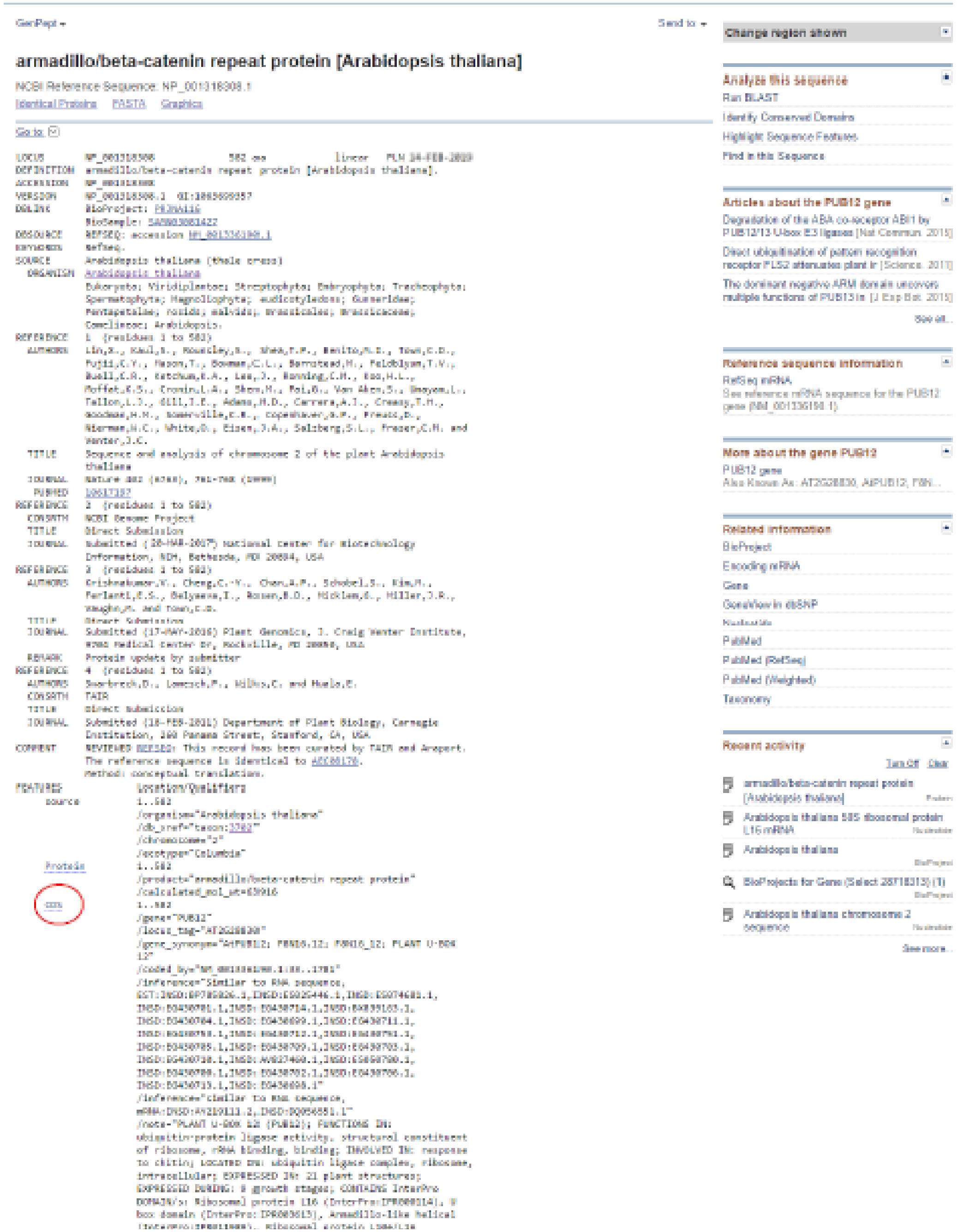
○ Abbreviated view
● Customize

Basic Features
○ All features
● Gene, RNA, and CDS features only

Display options
□ Show sequence
□ Show reverse complement
□ Show gap features

Bước 7.

- Go back to the GenBank record and click on the CDS link, just above the actual sequence (circled in red in Figure 3 on the previous page).
- ***Where did this take you or what happened when you did this?***



GenPept ▾ Send to: ▾

armadillo/beta-catenin repeat protein [Arabidopsis thaliana]

NCBI Reference Sequence: NP_001318308.1

Identical Proteins FASTA Graphics

Go to: ☑

```
LOCUS       NP_001318308             582 aa            linear   PLN 24-FEB-2019
DEFINITION  armadillo/beta-catenin repeat protein [Arabidopsis thaliana].
ACCESSION   NP_001318308
VERSION     NP_001318308.1  GI:1063699357
DBLINK      BioProject: PRJNA116
            BioSample: SAMN03081427
DBSOURCE    REFSEQ: accession NM_001336198.1
KEYWORDS    RefSeq.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae;
            Pentapetalae; rosids; malvids; Brassicales; Brassicaceae;
            Camelineae; Arabidopsis.
REFERENCE   1  (residues 1 to 582)
  AUTHORS   Lin,X., Kaul,S., Rounsley,S., Shea,T.P., Benito,M.I., Town,C.D.,
            Fujii,C.Y., Mason,T., Bowman,C.L., Barnstead,M., Feldblyum,T.V.,
            Buell,C.R., Ketchum,K.A., Lee,J., Ronning,C.M., Koo,H.L.,
            Moffat,K.S., Cronin,L.A., Shen,M., Pai,G., Van Aken,S., Umayam,L.,
            Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H.,
            Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D.,
            Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and
            Venter,J.C.
  TITLE     Sequence and analysis of chromosome 2 of the plant Arabidopsis
            thaliana
  JOURNAL   Nature 402 (6763), 761-768 (1999)
   PUBMED   10617197
REFERENCE   2  (residues 1 to 582)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-2017) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 582)
  AUTHORS   Krishnakumar,V., Cheng,C.-Y., Chan,A.P., Schobel,S., Kim,M.,
            Ferlanti,E.S., Belyaeva,I., Rosen,B.D., Micklem,G., Miller,J.R.,
            Vaughn,M. and Town,C.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-2016) Plant Genomics, J. Craig Venter Institute,
            9704 Medical Center Dr, Rockville, MD 20850, USA
  REMARK    Protein update by submitter
REFERENCE   4  (residues 1 to 582)
  AUTHORS   Swarbreck,D., Lamesch,P., Wilks,C. and Huala,E.
  CONSRTM   TAIR
  TITLE     Direct Submission
  JOURNAL   Submitted (18-FEB-2011) Department of Plant Biology, Carnegie
            Institution, 260 Panama Street, Stanford, CA, USA
COMMENT     REVIEWED REFSEQ: This record has been curated by TAIR and Araport.
            The reference sequence is identical to AEC08178.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..582
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /ecotype="Columbia"
     Protein         1..582
                     /product="armadillo/beta-catenin repeat protein"
                     /calculated_mol_wt=63916
     CDS             1..582
                     /gene="PUB12"
                     /locus_tag="AT2G28830"
                     /gene_synonym="AtPUB12; F8N16.12; F8N16_12; PLANT U-BOX
                     12"
                     /coded_by="NM_001336198.1:[illegible]"
                     /inference="Similar to RNA sequence,
                     EST:INSD:EP785826.1,INSD:ES025446.1,INSD:ES074681.1,
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]"
                     /inference="similar to RNA sequence,
                     mRNA:INSD:AY219111.2,INSD:DQ056551.1"
                     /note="PLANT U-BOX 12 (PUB12); FUNCTIONS IN:
                     ubiquitin-protein ligase activity, structural constituent
                     of ribosome, rRNA binding, binding; INVOLVED IN: response
                     to chitin; LOCATED IN: ubiquitin ligase complex, ribosome,
                     intracellular; EXPRESSED IN: 21 plant structures;
                     EXPRESSED DURING: 8 growth stages; CONTAINS InterPro
                     DOMAIN/s: Ribosomal protein L16 (InterPro:IPR000114), U
                     box domain (InterPro:IPR003613), Armadillo-like helical
                     (InterPro:IPR011989), Ribosomal protein L10e/L16
```

Change region shown

Analyze this sequence

Run BLAST

Identify Conserved Domains

Highlight Sequence Features

Find in this Sequence

Articles about the PUB12 gene

Degradation of the ABA co-receptor ABI1 by PUB12/13 U-box E3 ligases [Nat Commun. 2015]

Direct ubiquitination of pattern recognition receptor FLS2 attenuates plant in [Science. 2011]

The dominant negative ARM domain uncovers multiple functions of PUB13 in [J Exp Bot. 2015]

See all...

Reference sequence information

RefSeq mRNA

See reference mRNA sequence for the PUB12 gene (NM_001336198.1).

More about the gene PUB12

PUB12 gene

Also Known As: AT2G28830, AtPUB12, F8N...

Related information

BioProject

Encoding mRNA

Gene

GeneView in dbSNP

Nucleotide

PubMed

PubMed (RefSeq)

PubMed (Weighted)

Taxonomy

Recent activity

Turn Off Clear

armadillo/beta-catenin repeat protein [Arabidopsis thaliana] Protein

Arabidopsis thaliana 50S ribosomal protein L16 mRNA Nucleotide

Arabidopsis thaliana BioProject

BioProjects for Gene (Select 28718313) (1) BioProject

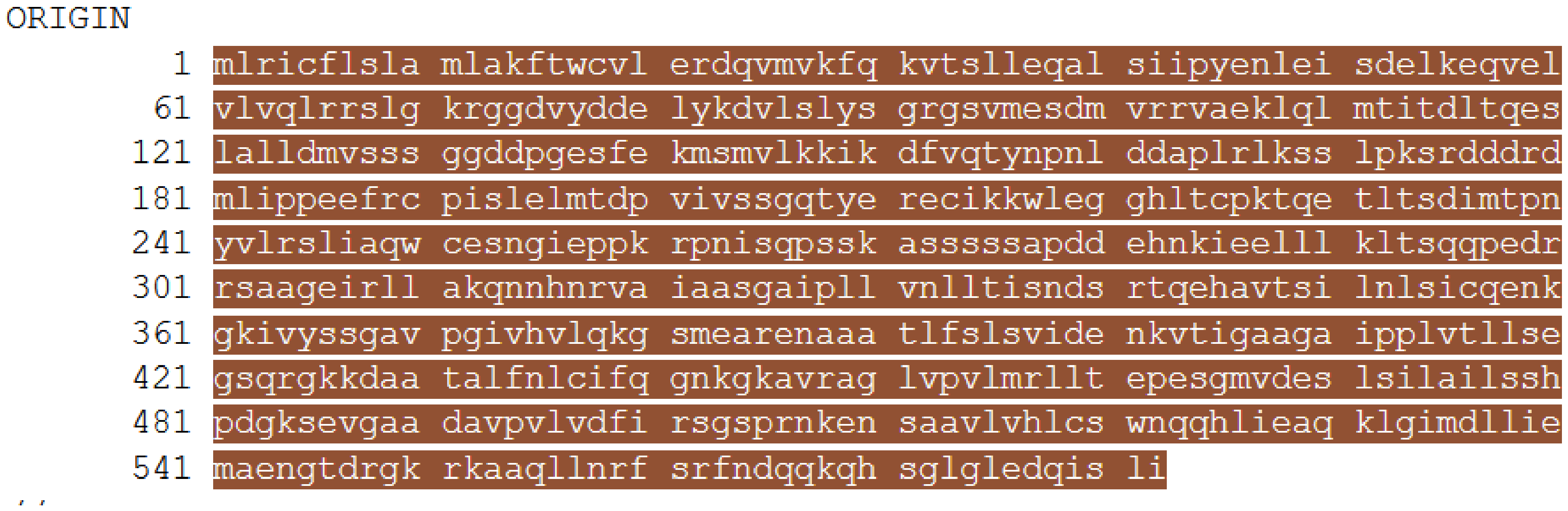
Arabidopsis thaliana chromosome 2 sequence Nucleotide

See more...

https://www.ncbi.nlm.nih.gov/protein/NP_001318308.1/

CDS : coding sequence

- The coding region of a gene, also known as the CDS (from coding sequence), is the portion of a gene's DNA or RNA that codes for protein.

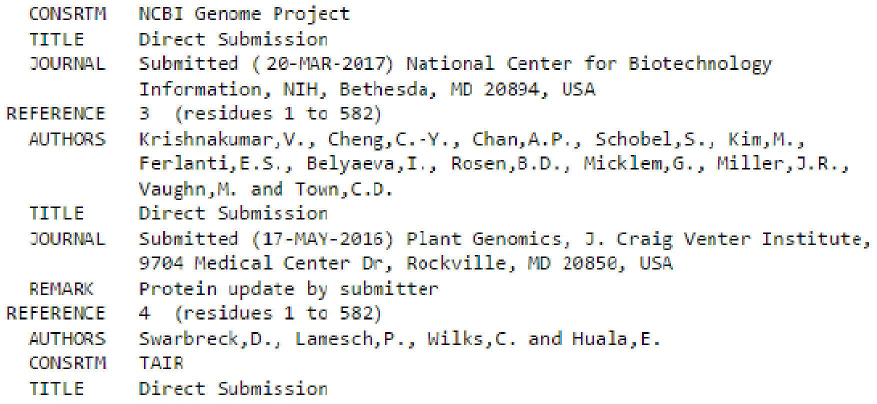


```
ORIGIN
        1 mlricflsla mlakftwcvl erdqvmvkfq kvtslleqal siipyenlei sdelkeqvel
       61 vlvqlrrslg krggdvydde lykdvlslys grgsvmesdm vrrvaeklql mtitdltqes
      121 lalldmvsss ggddpgesfe kmsmvlkkik dfvqtynpnl ddaplrlkss lpksrdddrd
      181 mlippeefrc pislelmtdp vivssgqtye recikkwleg ghltcpktqe tltsdimtpn
      241 yvlrsliaqw cesngieppk rpnisqpssk asssssapdd ehnkieelll kltsqqpedr
      301 rsaageirll akqnnhnrva iaasgaipll vnlltisnds rtqehavtsi lnlsicqenk
      361 gkivyssgav pgivhvlqkg smearenaaa tlfslsvide nkvtigaaga ipplvtllse
      421 gsqrgkkdaa talfnlcifq gnkgkavrag lvpvlmrllt epesgmvdes lsilailssh
      481 pdgksevgaa davpvlvdfi rsgsprnken saavlvhlcs wnqqhlieaq klgimdllie
      541 maengtdrgk rkaaqllnrf srfndqqkqh sglgledqis li
//
```

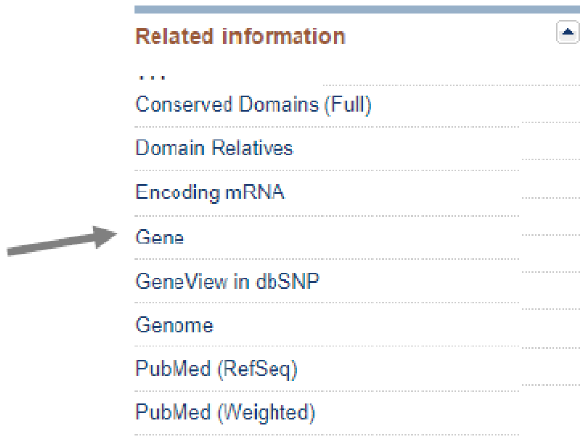
https://www.ncbi.nlm.nih.gov/protein/NP_001318308.1/

Bước 8.

- Go back to the GenBank record and examine the ***Related Information section*** on the lower right.
- This gives you direct links to other databases with information on this query. Find the Gene link.



```
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted ( 20-MAR-2017) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 582)
  AUTHORS   Krishnakumar,V., Cheng,C.-Y., Chan,A.P., Schobel,S., Kim,M.,
            Ferlanti,E.S., Belyaeva,I., Rosen,B.D., Micklem,G., Miller,J.R.,
            Vaughn,M. and Town,C.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-2016) Plant Genomics, J. Craig Venter Institute,
            9704 Medical Center Dr, Rockville, MD 20850, USA
  REMARK    Protein update by submitter
REFERENCE   4  (residues 1 to 582)
  AUTHORS   Swarbreck,D., Lamesch,P., Wilks,C. and Huala,E.
  CONSRTM   TAIR
  TITLE     Direct Submission
```



Related information

...

Conserved Domains (Full)

Domain Relatives

Encoding mRNA

Gene

GeneView in dbSNP

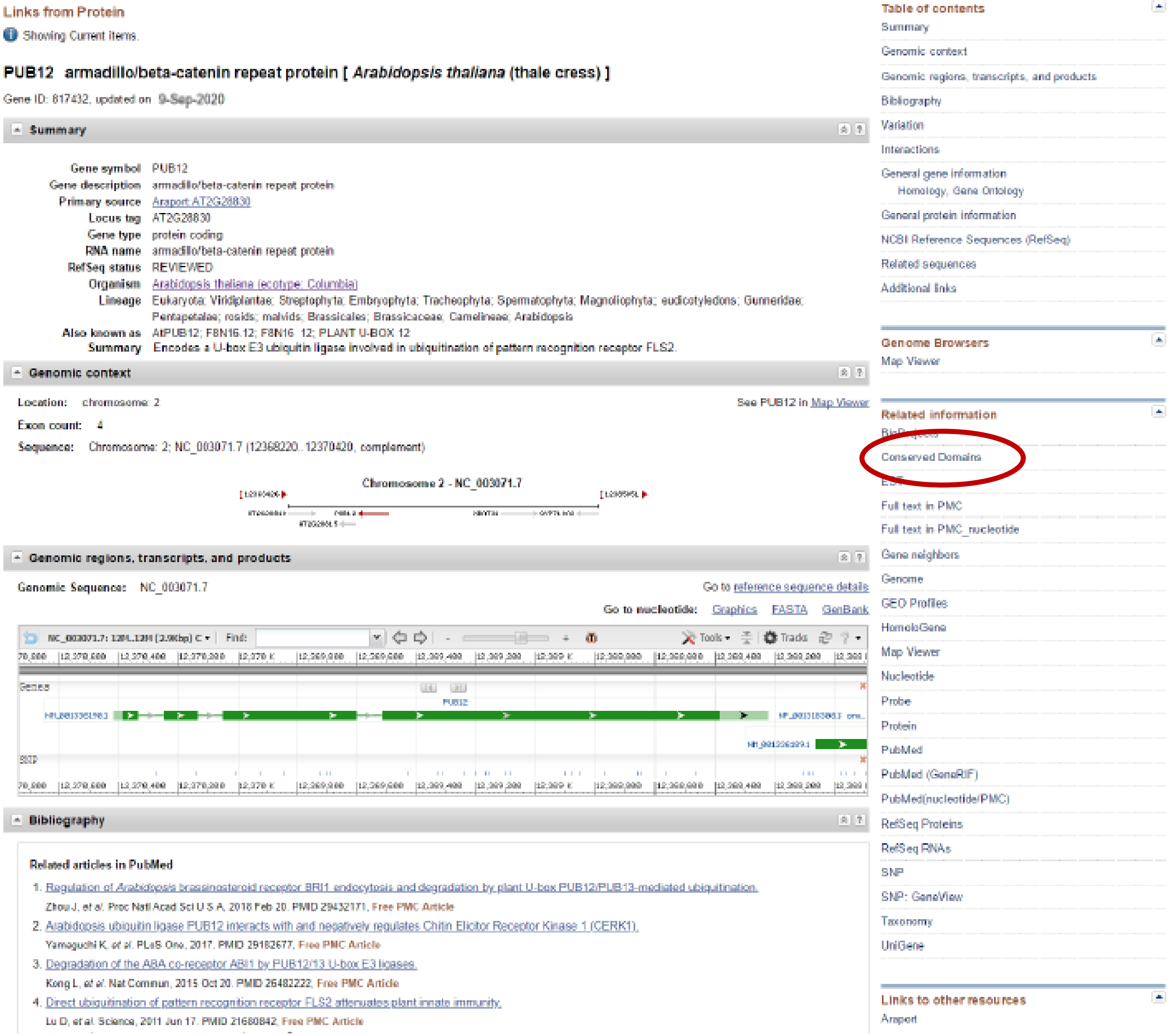
Genome

PubMed (RefSeq)

PubMed (Weighted)

https://www.ncbi.nlm.nih.gov/protein/NP_001318308.1/

Truncated GenBank Gene page for At2g28830 (also known as PUB12), the gene that encodes NP_001318308 protein



Links from Protein

Showing Current items.

PUB12 armadillo/beta-catenin repeat protein [*Arabidopsis thaliana* (thale cress)]

Gene ID: 817432, updated on 9-Sep-2020

Summary

Gene symbol	PUB12
Gene description	armadillo/beta-catenin repeat protein
Primary source	Araport:AT2G28830
Locus tag	AT2G28830
Gene type	protein coding
RNA name	armadillo/beta-catenin repeat protein
RefSeq status	REVIEWED
Organism	Arabidopsis thaliana (ecotype: Columbia)
Lineage	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae; Pentapetalae; rosids; malvids; Brassicales; Brassicaceae; Camelineae; Arabidopsis
Also known as	AtPUB12; F8N16.12; F8N16_12; PLANT U-BOX 12
Summary	Encodes a U-box E3 ubiquitin ligase involved in ubiquitination of pattern recognition receptor FLS2.

Genomic context

Location: chromosome: 2

See PUB12 in Map Viewer

Exon count: 4

Sequence: Chromosome: 2; NC_003071.7 (12368220..12370420, complement)

Genomic regions, transcripts, and products

Genomic Sequence: NC_003071.7

Go to reference sequence details

Go to nucleotide: Graphics FASTA GenBank

Bibliography

Related articles in PubMed

1. Regulation of *Arabidopsis* brassinosteroid receptor BRI1 endocytosis and degradation by plant U-box PUB12/PUB13-mediated ubiquitination.
   Zhou J, *et al.* Proc Natl Acad Sci U S A, 2018 Feb 20. PMID 29432171, Free PMC Article
2. Arabidopsis ubiquitin ligase PUB12 interacts with and negatively regulates Chitin Elicitor Receptor Kinase 1 (CERK1).
   Yamaguchi K, *et al.* PLoS One, 2017. PMID 29182677, Free PMC Article
3. Degradation of the ABA co-receptor ABI1 by PUB12/13 U-box E3 ligases.
   Kong L, *et al.* Nat Commun, 2015 Oct 20. PMID 26482222, Free PMC Article
4. Direct ubiquitination of pattern recognition receptor FLS2 attenuates plant innate immunity.
   Lu D, *et al.* Science, 2011 Jun 17. PMID 21680842, Free PMC Article

Genome Browsers

Map Viewer

Related information

BioProjects

Conserved Domains

EST

Full text in PMC

Full text in PMC_nucleotide

Gene neighbors

Genome

GEO Profiles

HomoloGene

Map Viewer

Nucleotide

Probe

Protein

PubMed

PubMed (GeneRIF)

PubMed(nucleotide/PMC)

RefSeq Proteins

RefSeq RNAs

SNP

SNP: GeneView

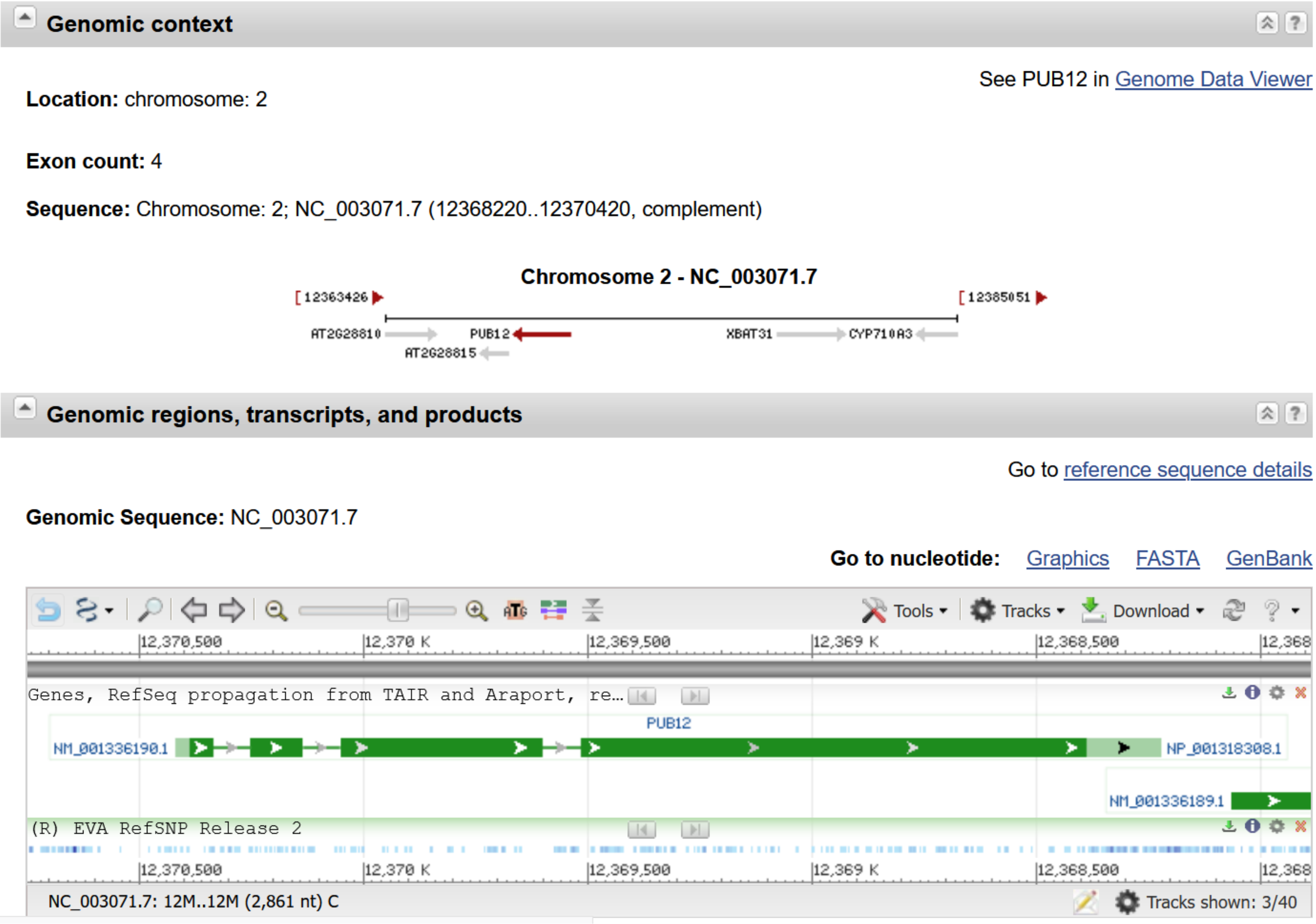
Taxonomy

UniGene

Links to other resources

Araport

https://www.ncbi.nlm.nih.gov/gene?LinkName=protein_gene&from_uid=1063699357



Genomic context

See PUB12 in Genome Data Viewer

Location: chromosome: 2

Exon count: 4

Sequence: Chromosome: 2; NC_003071.7 (12368220..12370420, complement)

Genomic regions, transcripts, and products

Go to reference sequence details

Genomic Sequence: NC_003071.7

Go to nucleotide: Graphics FASTA GenBank

NC_003071.7: 12M..12M (2,861 nt) C

Tracks shown: 3/40

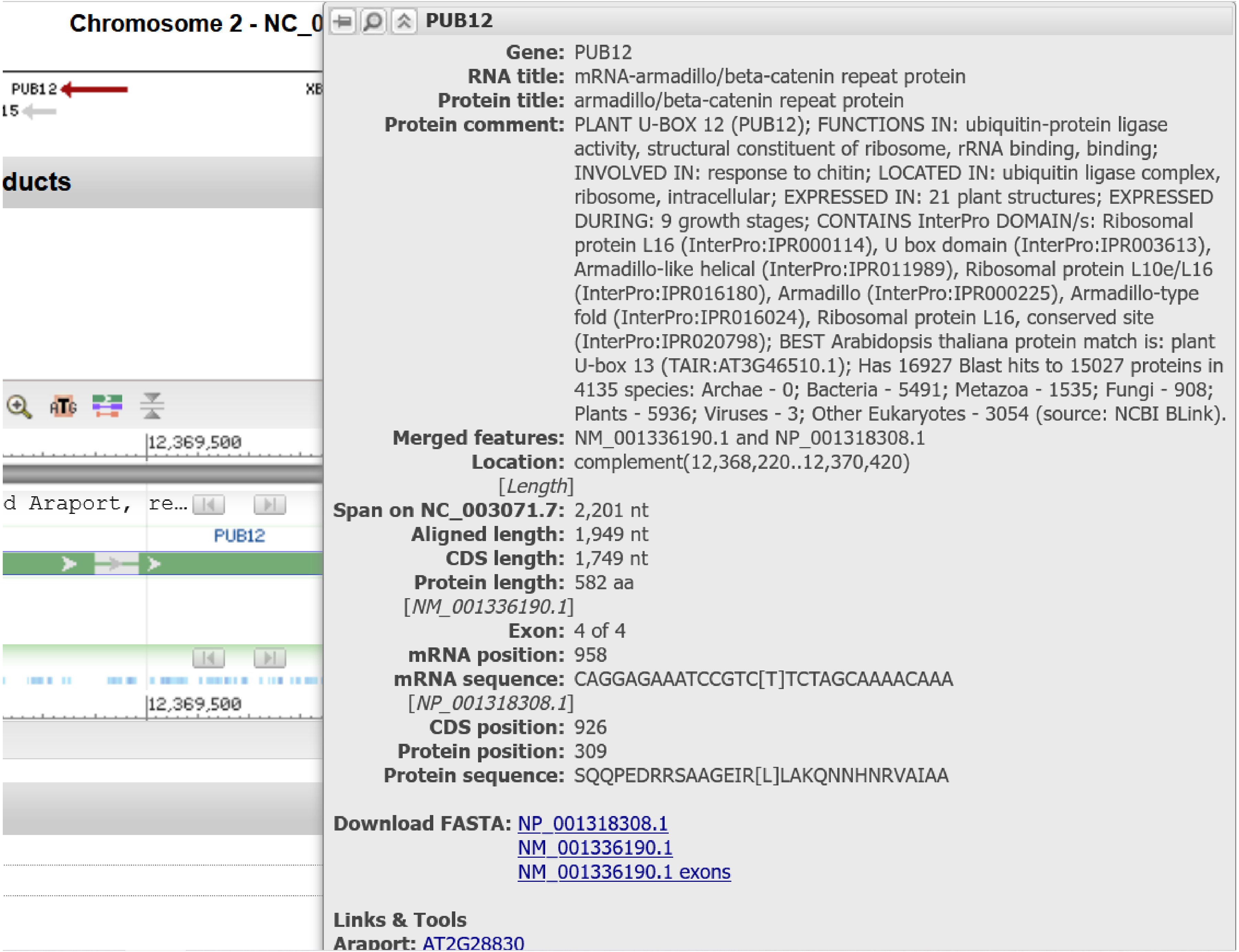
https://www.ncbi.nlm.nih.gov/gene?LinkName=protein_gene&from_uid=1063699357

Bước 9. Trả lời các câu hỏi sau

a. Where is your gene's location in the genome? (Tip: hover with your cursor over the green bars in the "Genomic regions, transcripts, and products" section; the green bars represent the gene in the sequence viewer)

b. How many exons do you see in this gene? Tip: how many green boxes are there?

Xác định các vùng introns trong gen này.



PUB12

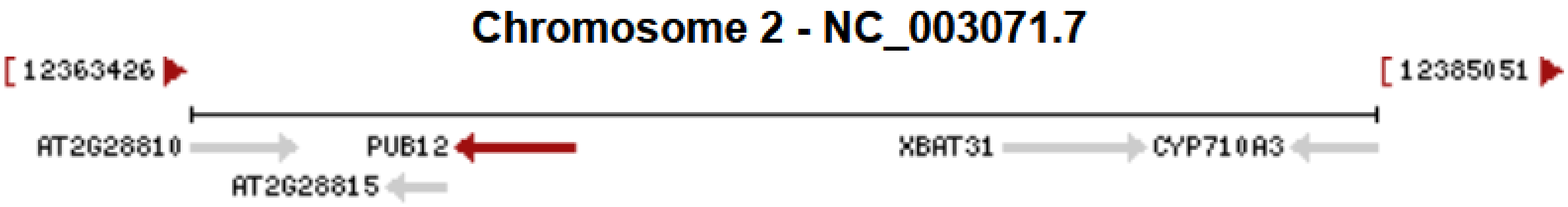
Gene: PUB12
RNA title: mRNA-armadillo/beta-catenin repeat protein
Protein title: armadillo/beta-catenin repeat protein
Protein comment: PLANT U-BOX 12 (PUB12); FUNCTIONS IN: ubiquitin-protein ligase activity, structural constituent of ribosome, rRNA binding, binding; INVOLVED IN: response to chitin; LOCATED IN: ubiquitin ligase complex, ribosome, intracellular; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 9 growth stages; CONTAINS InterPro DOMAIN/s: Ribosomal protein L16 (InterPro:IPR000114), U box domain (InterPro:IPR003613), Armadillo-like helical (InterPro:IPR011989), Ribosomal protein L10e/L16 (InterPro:IPR016180), Armadillo (InterPro:IPR000225), Armadillo-type fold (InterPro:IPR016024), Ribosomal protein L16, conserved site (InterPro:IPR020798); BEST Arabidopsis thaliana protein match is: plant U-box 13 (TAIR:AT3G46510.1); Has 16927 Blast hits to 15027 proteins in 4135 species: Archae - 0; Bacteria - 5491; Metazoa - 1535; Fungi - 908; Plants - 5936; Viruses - 3; Other Eukaryotes - 3054 (source: NCBI BLink).
Merged features: NM_001336190.1 and NP_001318308.1
Location: complement(12,368,220..12,370,420)
[*Length*]
Span on NC_003071.7: 2,201 nt
Aligned length: 1,949 nt
CDS length: 1,749 nt
Protein length: 582 aa
[*NM_001336190.1*]
Exon: 4 of 4
mRNA position: 958
mRNA sequence: CAGGAGAAATCCGTC[T]TCTAGCAAAACAAA
[*NP_001318308.1*]
CDS position: 926
Protein position: 309
Protein sequence: SQQPEDRRSAAGEIR[L]LAKQNNHNRVAIAA

Download FASTA: NP_001318308.1
NM_001336190.1
NM_001336190.1 exons

Links & Tools
Araport: AT2G28830

c. What are the names of the genes surrounding it (i.e. what is its “Genomic context”)?

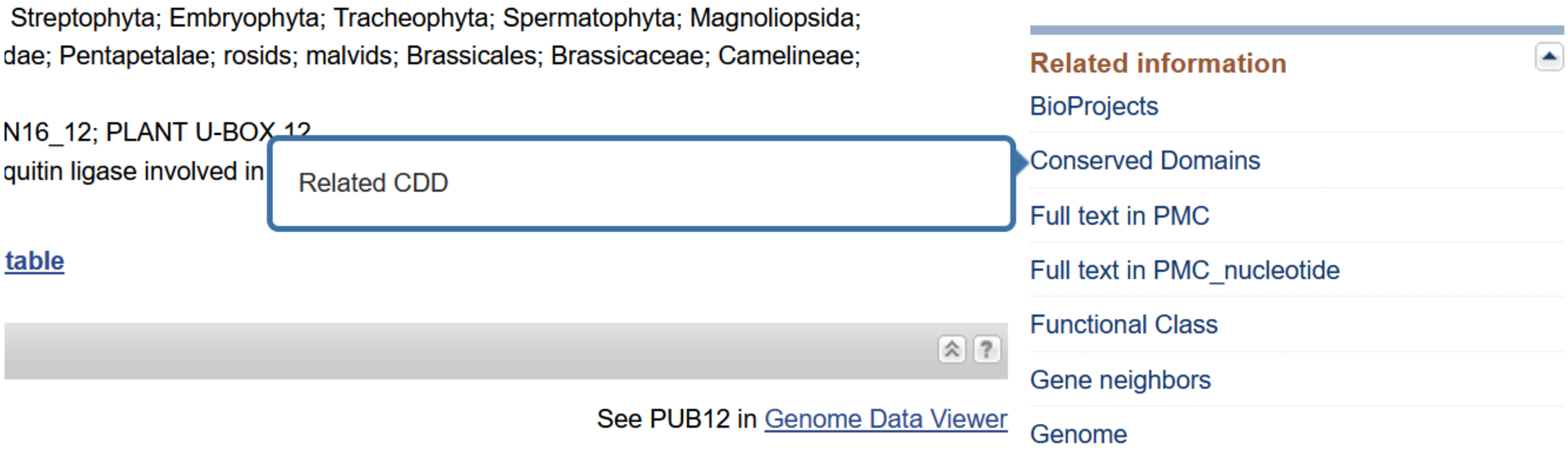
- **Location:** chromosome: 2
- **Exon count:** 4
- **Sequence:** Chromosome: 2; NC_003071.7 (12368220..12370420, complement)



https://www.ncbi.nlm.nih.gov/gene?LinkName=protein_gene&from_uid=1063699357

d. Does it have any conserved domains? What are they called?

(Tip: use the “Related Information” link to Conserved Domains on the right of the Gene page)



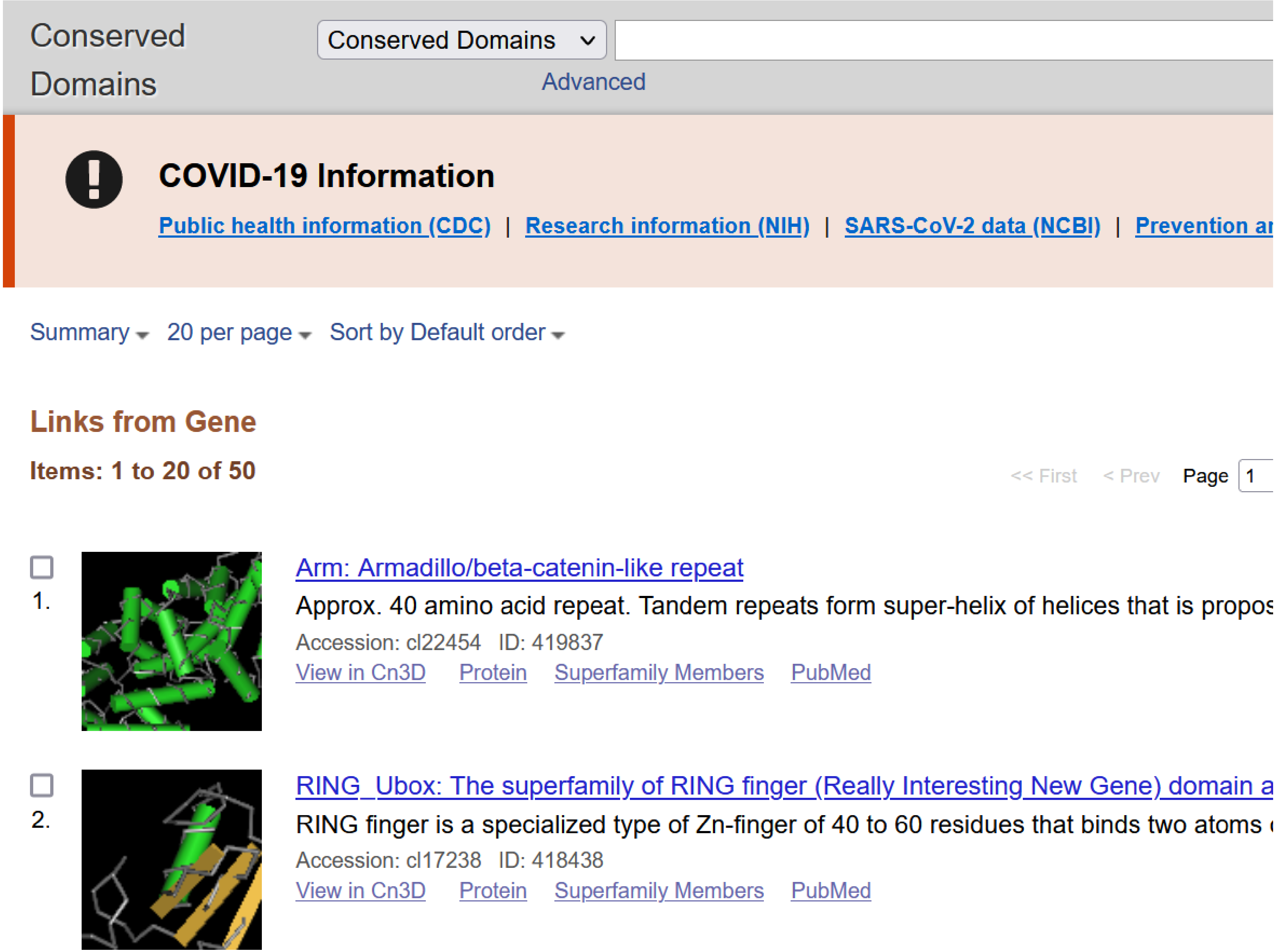
Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida;
dae; Pentapetalae; rosids; malvids; Brassicales; Brassicaceae; Camelineae;

N16_12; PLANT U-BOX 12
quitin ligase involved in

Related CDD

table

See PUB12 in Genome Data Viewer



Conserved Domains

Conserved Domains ⌄

Advanced

COVID-19 Information

Public health information (CDC) | Research information (NIH) | SARS-CoV-2 data (NCBI) | Prevention ar

Summary ▾ 20 per page ▾ Sort by Default order ▾

Links from Gene

Items: 1 to 20 of 50

<< First < Prev Page 1

□ 1.

Arm: Armadillo/beta-catenin-like repeat

Approx. 40 amino acid repeat. Tandem repeats form super-helix of helices that is propos

Accession: cl22454 ID: 419837

View in Cn3D Protein Superfamily Members PubMed

□ 2.

RING_Ubox: The superfamily of RING finger (Really Interesting New Gene) domain a

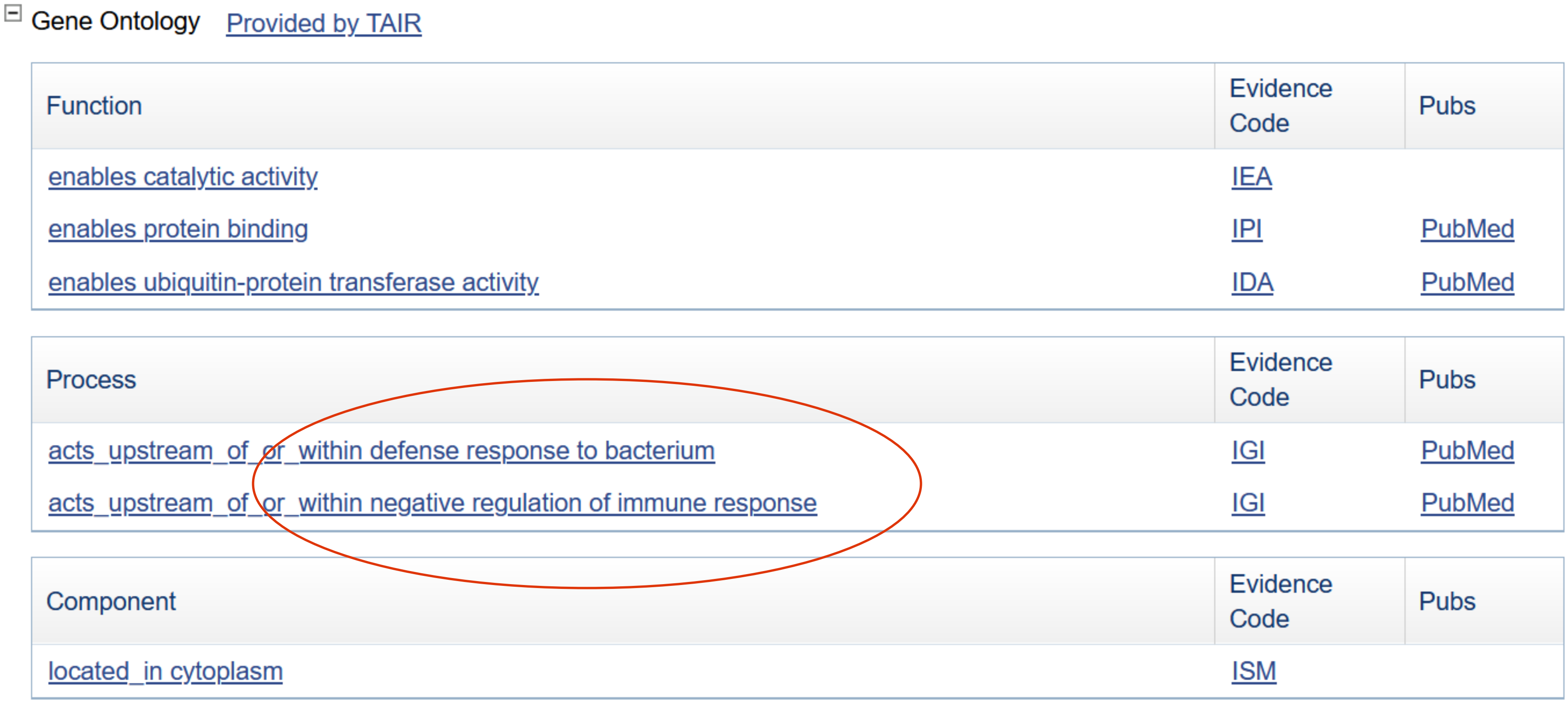
RING finger is a specialized type of Zn-finger of 40 to 60 residues that binds two atoms

Accession: cl17238 ID: 418438

View in Cn3D Protein Superfamily Members PubMed

https://www.ncbi.nlm.nih.gov/cdd?LinkName=gene_cdd&from_uid=817432

e. After exploring conserved domains go back to the Gene page. What biological process (Gene Ontology terms) is this gene involved with (scroll down!)?



⊟ Gene Ontology Provided by TAIR

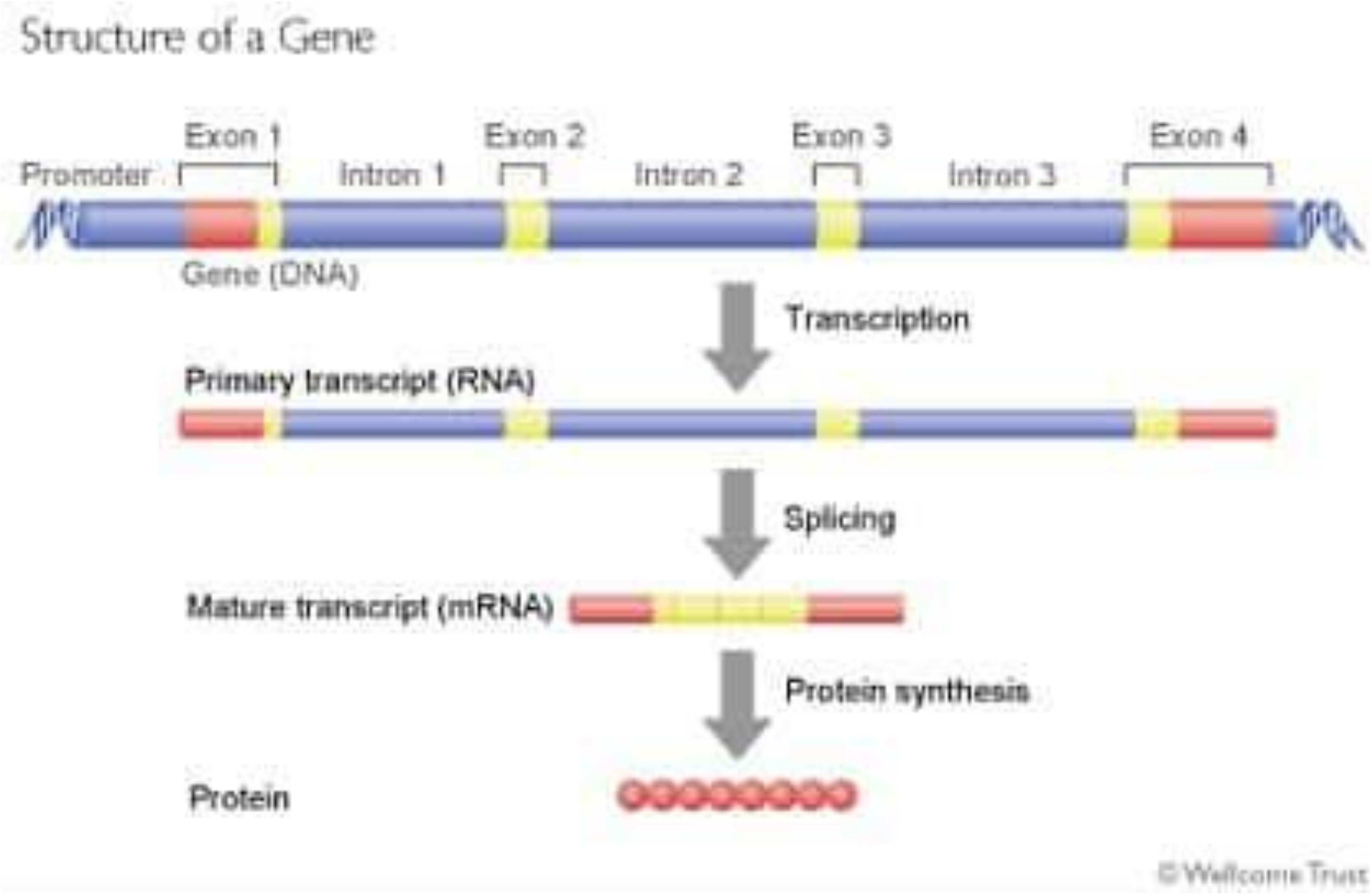
Function	Evidence Code	Pubs
enables catalytic activity	IEA	
enables protein binding	IPI	PubMed
enables ubiquitin-protein transferase activity	IDA	PubMed

Process	Evidence Code	Pubs
acts_upstream_of_or_within defense response to bacterium	IGI	PubMed
acts_upstream_of_or_within negative regulation of immune response	IGI	PubMed

Component	Evidence Code	Pubs
located_in cytoplasm	ISM	

https://www.ncbi.nlm.nih.gov/gene?LinkName=protein_gene&from_uid=1063699357

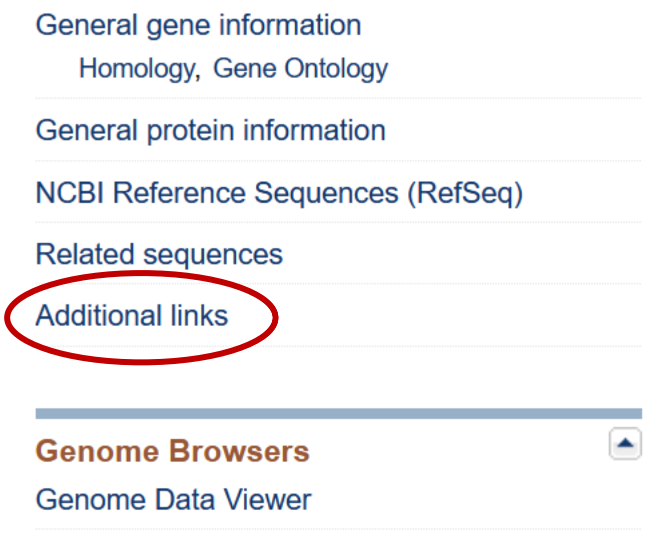
Structure of a Gene



Bước 10.

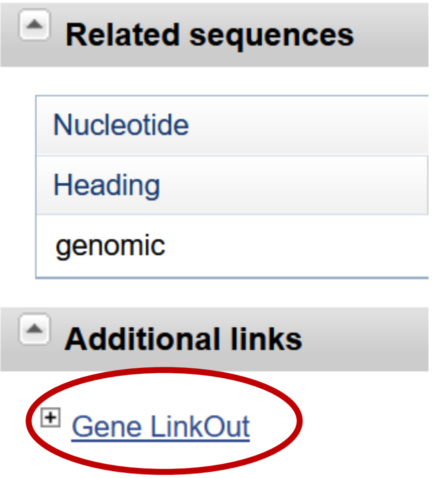
- On the Gene page, there are also Additional links to examine a gene’s structure, function and phylogenetic relationships further.
- The navigation sidebar on the right has an “Additional links” hyperlink which will take you to the bottom of the page, where they’re found for most genes.
- Click [+] Gene LinkOut to see them.

a. ***Click on Additional Links. What kind of information is in this section?***



General gene information
Homology, Gene Ontology
General protein information
NCBI Reference Sequences (RefSeq)
Related sequences
Additional links

Genome Browsers
Genome Data Viewer



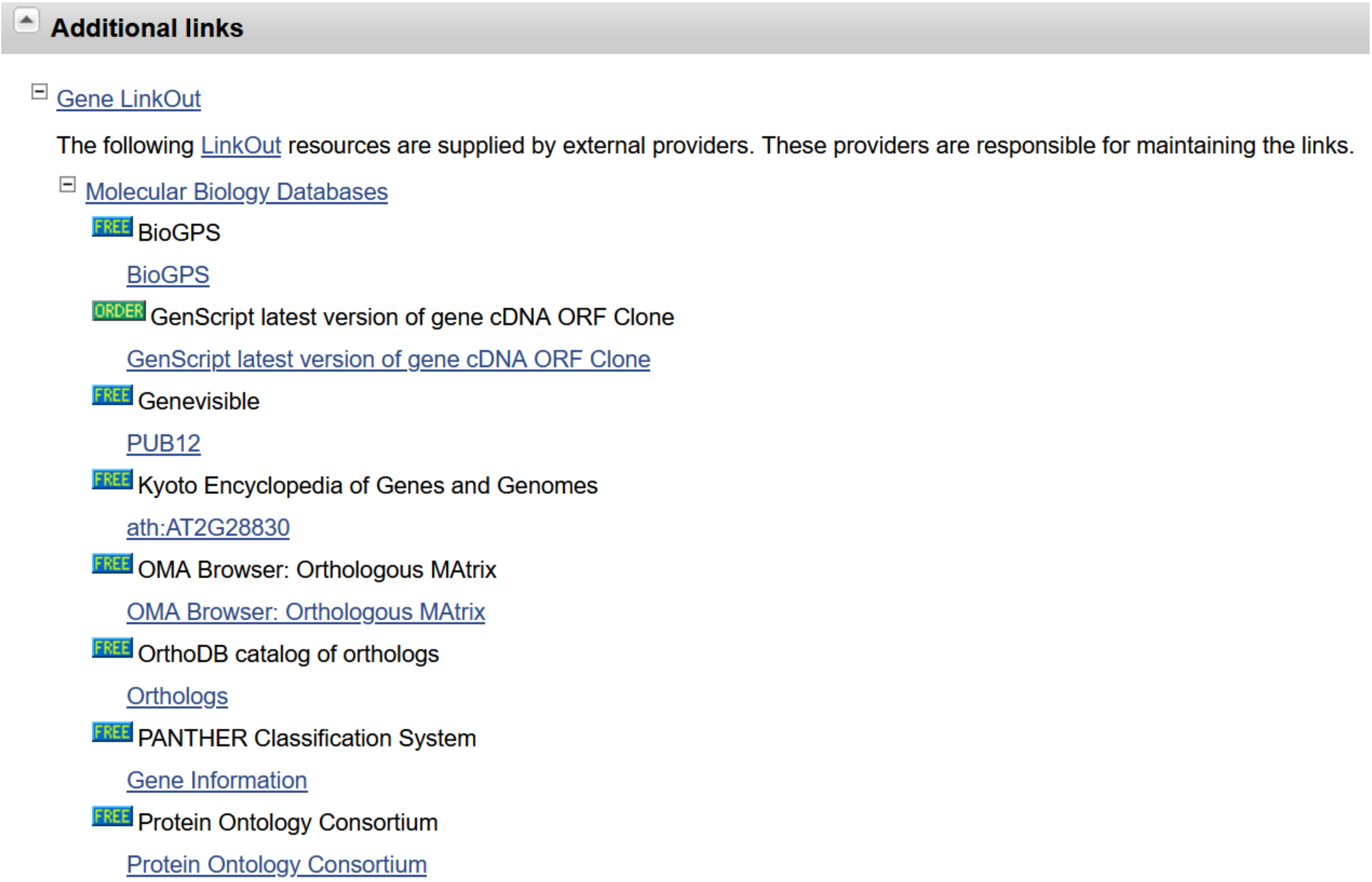
Related sequences

Nucleotide
Heading
genomic

Additional links

⊞ Gene LinkOut

https://www.ncbi.nlm.nih.gov/gene?LinkName=protein_gene&from_uid=1063699357



Additional links

⊟ Gene LinkOut

The following LinkOut resources are supplied by external providers. These providers are responsible for maintaining the links.

⊟ Molecular Biology Databases

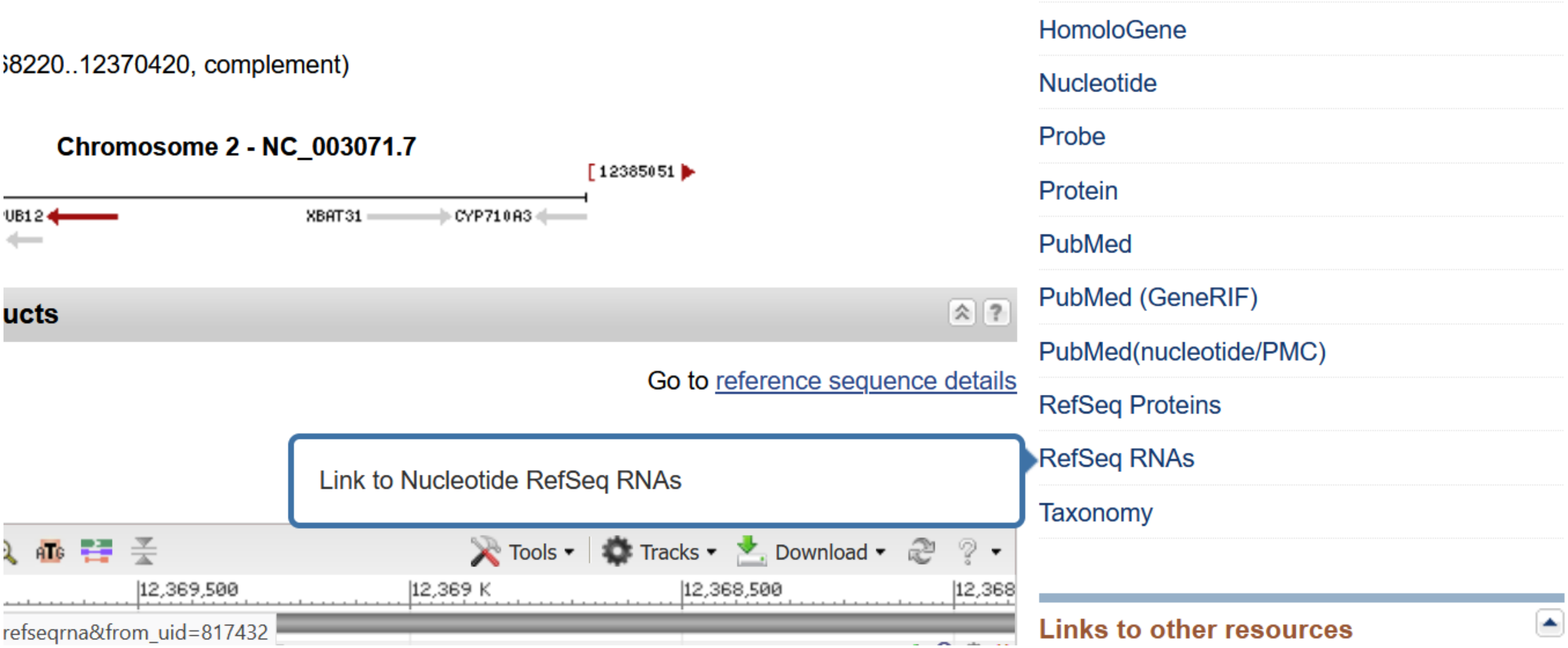
- FREE BioGPS
  - BioGPS
- ORDER GenScript latest version of gene cDNA ORF Clone
  - GenScript latest version of gene cDNA ORF Clone
- FREE Genevisible
  - PUB12
- FREE Kyoto Encyclopedia of Genes and Genomes
  - ath:AT2G28830
- FREE OMA Browser: Orthologous MAtrix
  - OMA Browser: Orthologous MAtrix
- FREE OrthoDB catalog of orthologs
  - Orthologs
- FREE PANTHER Classification System
  - Gene Information
- FREE Protein Ontology Consortium
  - Protein Ontology Consortium

https://www.ncbi.nlm.nih.gov/gene?LinkName=protein_gene&from_uid=1063699357#additional-links

Câu hỏi 10.b

- ***10.b. Why is the length of the mRNA different from the value you can calculate from the start and stop positions in Question 9a?***

Chọn mục RefSeq RNAs



https://www.ncbi.nlm.nih.gov/gene?LinkName=protein_gene&from_uid=1063699357

- Chọn mục RefSeq RNAs

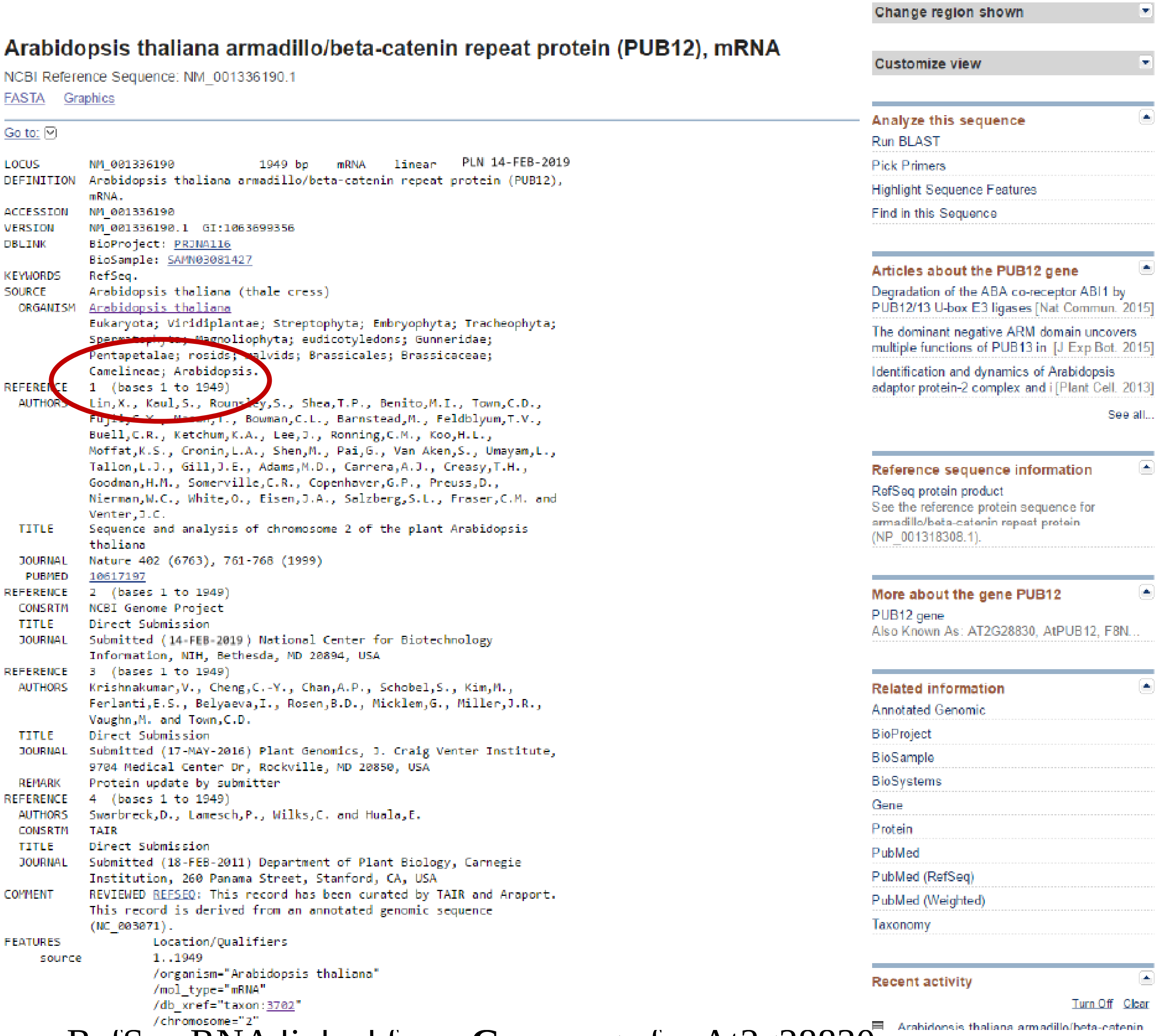
PubMed (GeneRIF)

PubMed(nucleotide/PMC)

RefSeq Proteins

RefSeq RNAs

Taxonomy



Arabidopsis thaliana armadillo/beta-catenin repeat protein (PUB12), mRNA

NCBI Reference Sequence: NM_001336190.1

FASTA Graphics

Go to:

```
LOCUS       NM_001336190            1949 bp    mRNA    linear   PLN 14-FEB-2019
DEFINITION  Arabidopsis thaliana armadillo/beta-catenin repeat protein (PUB12),
            mRNA.
ACCESSION   NM_001336190
VERSION     NM_001336190.1  GI:1063699356
DBLINK      BioProject: PRJNA116
            BioSample: SAMN03081427
KEYWORDS    RefSeq.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae;
            Pentapetalae; rosids; malvids; Brassicales; Brassicaceae;
            Camelineae; Arabidopsis.
REFERENCE   1  (bases 1 to 1949)
  AUTHORS   Lin,X., Kaul,S., Rounsley,S., Shea,T.P., Benito,M.I., Town,C.D.,
            Fujii,C.Y., Mason,T., Bowman,C.L., Barnstead,M., Feldblyum,T.V.,
            Buell,C.R., Ketchum,K.A., Lee,J., Ronning,C.M., Koo,H.L.,
            Moffat,K.S., Cronin,L.A., Shen,M., Pai,G., Van Aken,S., Umayam,L.,
            Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H.,
            Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D.,
            Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and
            Venter,J.C.
  TITLE     Sequence and analysis of chromosome 2 of the plant Arabidopsis
            thaliana
  JOURNAL   Nature 402 (6763), 761-768 (1999)
   PUBMED   10617197
REFERENCE   2  (bases 1 to 1949)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2019) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (bases 1 to 1949)
  AUTHORS   Krishnakumar,V., Cheng,C.-Y., Chan,A.P., Schobel,S., Kim,M.,
            Ferlanti,E.S., Belyaeva,I., Rosen,B.D., Micklem,G., Miller,J.R.,
            Vaughn,M. and Town,C.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-2016) Plant Genomics, J. Craig Venter Institute,
            9704 Medical Center Dr, Rockville, MD 20850, USA
  REMARK    Protein update by submitter
REFERENCE   4  (bases 1 to 1949)
  AUTHORS   Swarbreck,D., Lamesch,P., Wilks,C. and Huala,E.
  CONSRTM   TAIR
  TITLE     Direct Submission
  JOURNAL   Submitted (18-FEB-2011) Department of Plant Biology, Carnegie
            Institution, 260 Panama Street, Stanford, CA, USA
COMMENT     REVIEWED REFSEQ: This record has been curated by TAIR and Araport.
            This record is derived from an annotated genomic sequence
            (NC_003071).
FEATURES             Location/Qualifiers
     source          1..1949
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="2"
```

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Articles about the PUB12 gene

Degradation of the ABA co-receptor ABI1 by PUB12/13 U-box E3 ligases [Nat Commun. 2015]

The dominant negative ARM domain uncovers multiple functions of PUB13 in [J Exp Bot. 2015]

Identification and dynamics of Arabidopsis adaptor protein-2 complex and i [Plant Cell. 2013]

See all...

Reference sequence information

RefSeq protein product

See the reference protein sequence for armadillo/beta-catenin repeat protein (NP_001318308.1).

More about the gene PUB12

PUB12 gene

Also Known As: AT2G28830, AtPUB12, F8N...

Related information

Annotated Genomic

BioProject

BioSample

BioSystems

Gene

Protein

PubMed

PubMed (RefSeq)

PubMed (Weighted)

Taxonomy

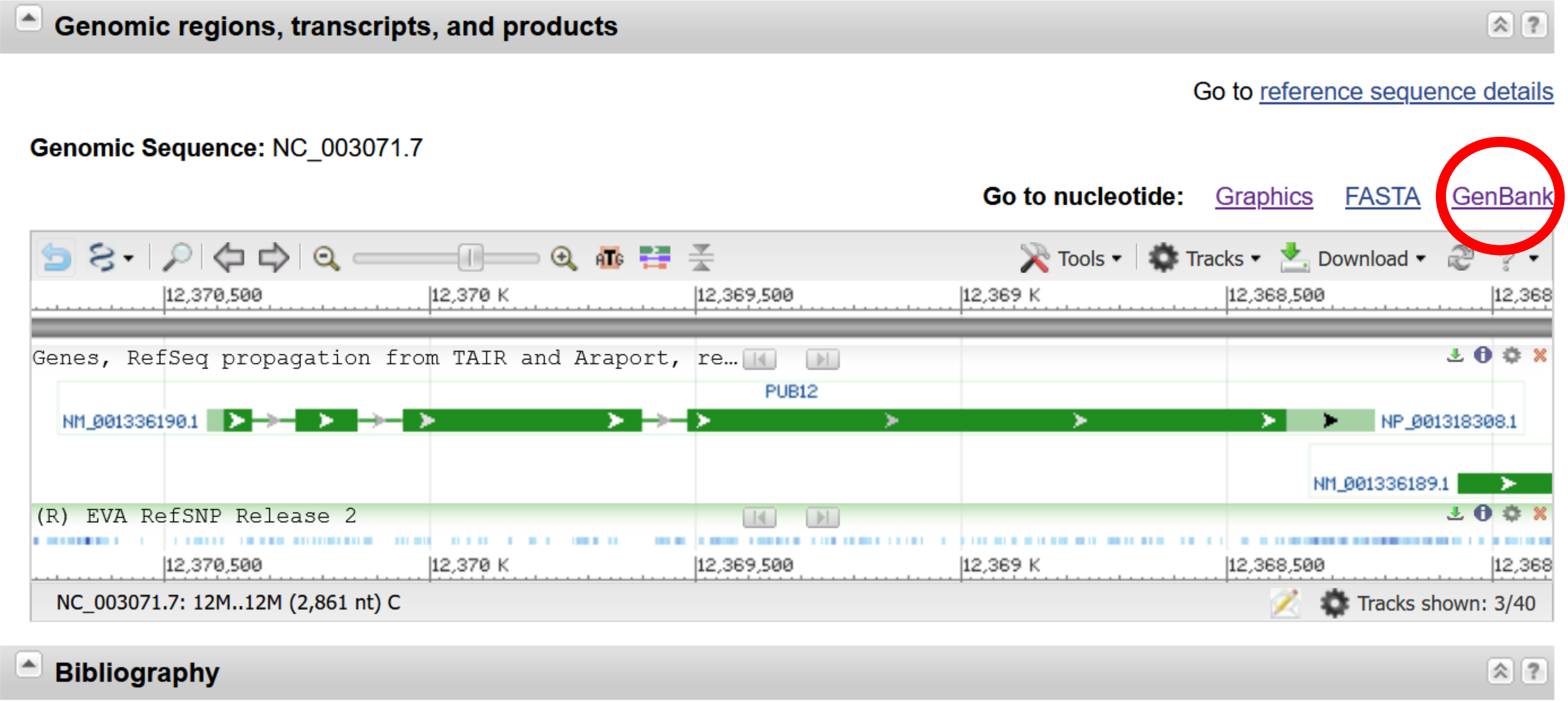
Recent activity

Arabidopsis thaliana armadillo/beta-catenin

RefSeq RNA linked from **Gene** page for At2g28830

https://www.ncbi.nlm.nih.gov/nuccore/1063699356

Xem thông tin exons, introns



Genomic regions, transcripts, and products

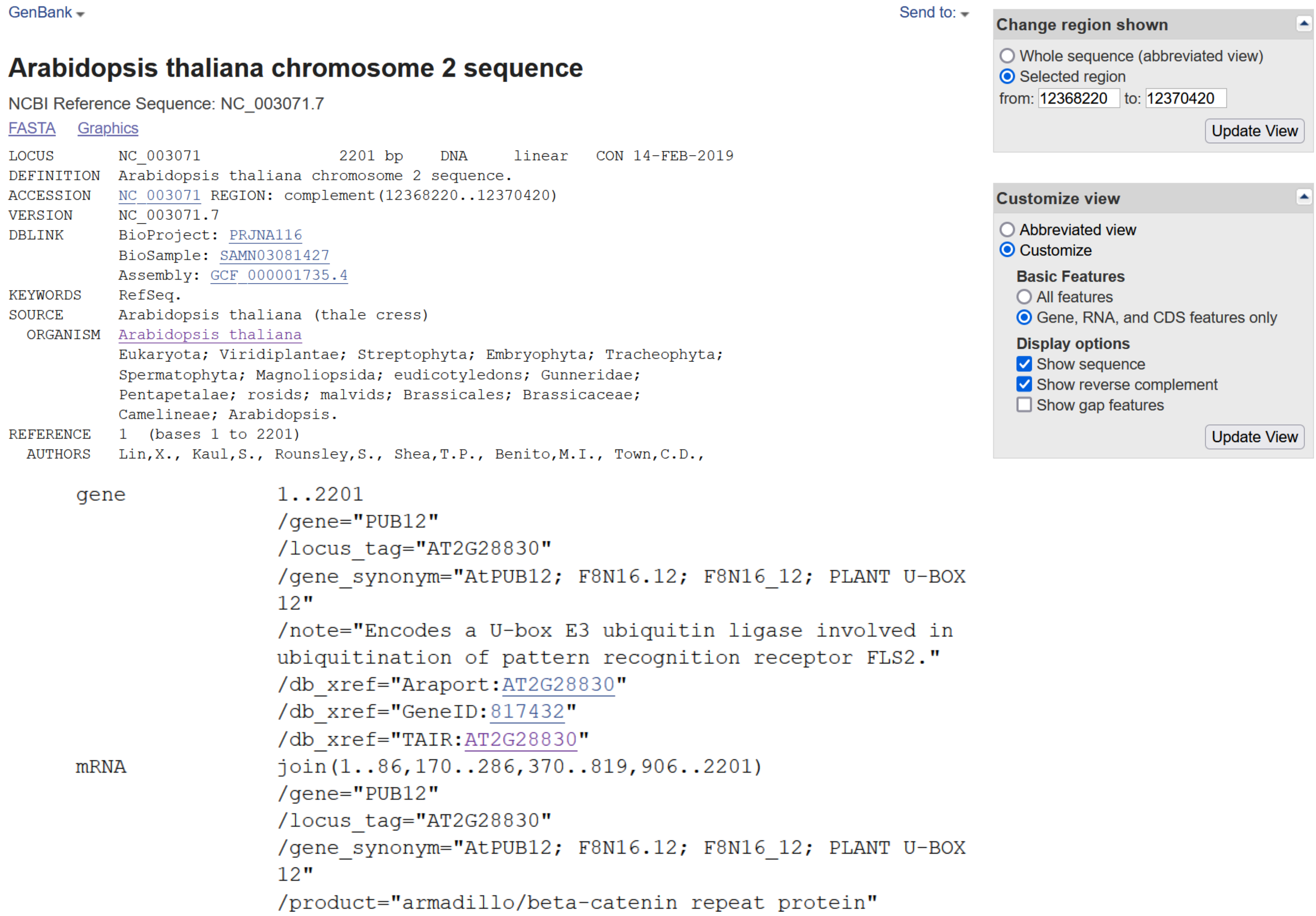
Go to reference sequence details

Genomic Sequence: NC_003071.7

Go to nucleotide: Graphics FASTA GenBank

Bibliography

https://www.ncbi.nlm.nih.gov/gene?LinkName=protein_gene&from_uid=1063699357



Arabidopsis thaliana chromosome 2 sequence

NCBI Reference Sequence: NC_003071.7

```
LOCUS       NC_003071               2201 bp    DNA     linear   CON 14-FEB-2019
DEFINITION  Arabidopsis thaliana chromosome 2 sequence.
ACCESSION   NC_003071 REGION: complement(12368220..12370420)
VERSION     NC_003071.7
DBLINK      BioProject: PRJNA116
            BioSample: SAMN03081427
            Assembly: GCF_000001735.4
KEYWORDS    RefSeq.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliopsida; eudicotyledons; Gunneridae;
            Pentapetalae; rosids; malvids; Brassicales; Brassicaceae;
            Camelineae; Arabidopsis.
REFERENCE   1  (bases 1 to 2201)
  AUTHORS   Lin,X., Kaul,S., Rounsley,S., Shea,T.P., Benito,M.I., Town,C.D.,
```

Change region shown

○ Whole sequence (abbreviated view)
● Selected region
from: 12368220 to: 12370420

Update View

Customize view

○ Abbreviated view
● Customize

Basic Features

○ All features
● Gene, RNA, and CDS features only

Display options

☑ Show sequence
☑ Show reverse complement
□ Show gap features

Update View

```
     gene            1..2201
                     /gene="PUB12"
                     /locus_tag="AT2G28830"
                     /gene_synonym="AtPUB12; F8N16.12; F8N16_12; PLANT U-BOX
                     12"
                     /note="Encodes a U-box E3 ubiquitin ligase involved in
                     ubiquitination of pattern recognition receptor FLS2."
                     /db_xref="Araport:AT2G28830"
                     /db_xref="GeneID:817432"
                     /db_xref="TAIR:AT2G28830"
     mRNA            join(1..86,170..286,370..819,906..2201)
                     /gene="PUB12"
                     /locus_tag="AT2G28830"
                     /gene_synonym="AtPUB12; F8N16.12; F8N16_12; PLANT U-BOX
                     12"
                     /product="armadillo/beta-catenin repeat protein"
```

https://www.ncbi.nlm.nih.gov/nuccore/NC_003071.7?report=genbank&from=12368220&to=12370420&strand=true

Next: Lab 1b. Basic BLAST