

How to find and get sequence data



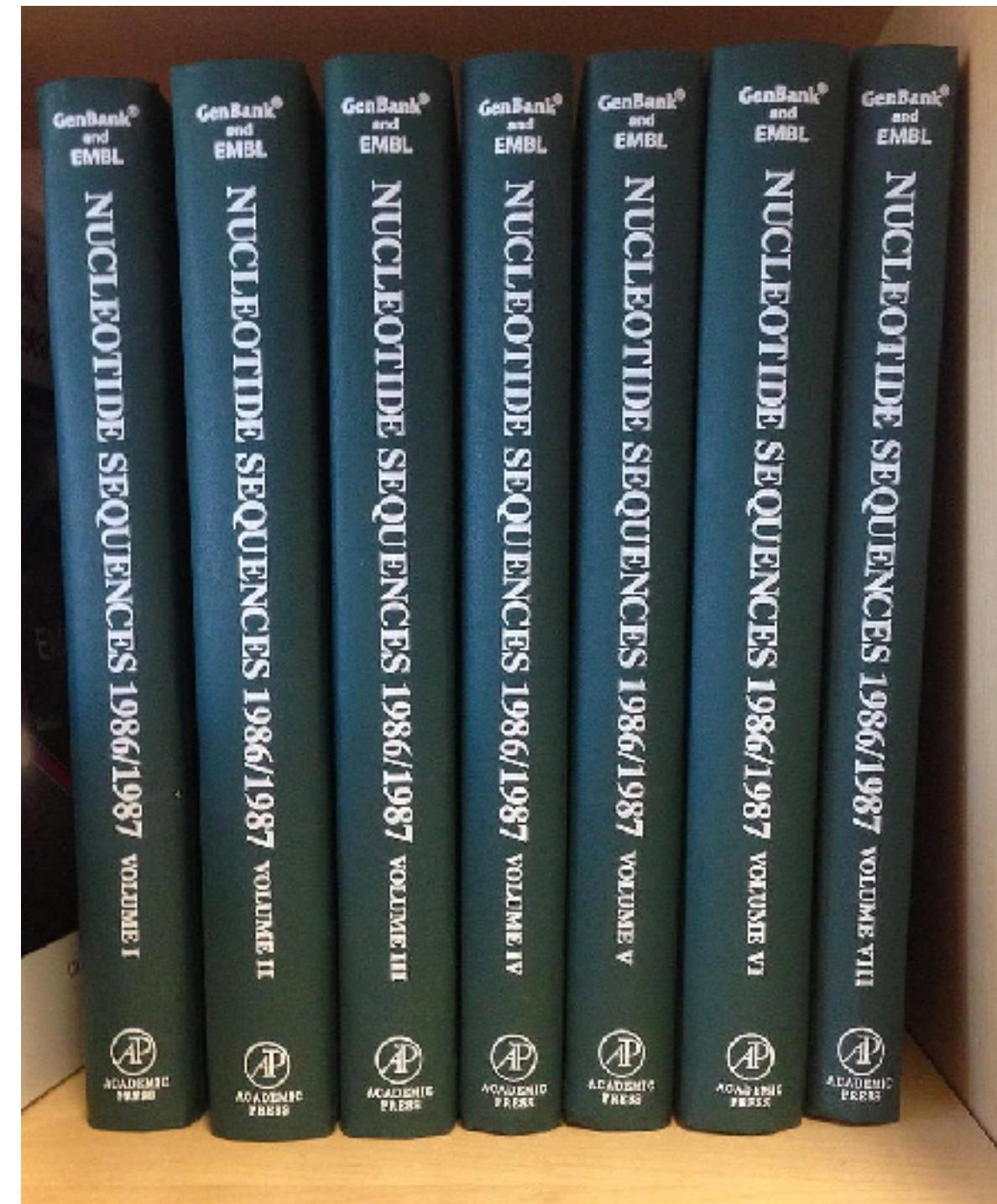
Computational Biology
Workshop

Todos Santos Center
May 9-12, 2022

Mark Stenglein, PhD
Associate Professor
Center for Vector-Borne Infectious Diseases
Department of Microbiology, Immunology, and Pathology
Mark.Stenglein@colostate.edu
StengleinLab.org

GenBank was one of the earliest sequence databases.

GenBank circa 1987



~10,000 sequences

GenBank release 100 (1997)
distributed by CDROM



Genbank today



>237,000,000 sequences

BOVCHYMOA NUCLEOTIDE SEQUENCES 1984										
<small>SITES: key site span description key site span description</small>										
refnumbr	21	1	numbered 1 in [1]	pept/pept	195	0	chymo propept end/ mature pept			
-pept	21	1	chymo prepropept cds start				start			
pept/pept	69	0	chymo prepropept end/ propept start pept<-	1166	1	chymo mature pept cds end				
ORIGIN:	20 bases upstream from codon 1									
SEQUENCES:	1275	bp	293	a	391	c	336	g	255	t
1	ccgtttttttt	caatggccaa	atggagggttc	tctgttttgt	tttgcgtttttt	ccccggggcc	tgatgttacc	aggatccctt	tgttacaaaa	gg
101	caatgtttttt	aggaggccgt	tgaaggagca	tgggttttttgc	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt
201	gttggccatcg	ttcccttgc	caatcttttgc	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt
301	gtttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt
401	ggggaaagcc	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt
501	gttgggtttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt
601	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt
701	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt
801	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt
901	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt
1001	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt
1101	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt
1201	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt	ttttttttttttt

~1,300,000 sequences

First release: 1982: 606 sequences

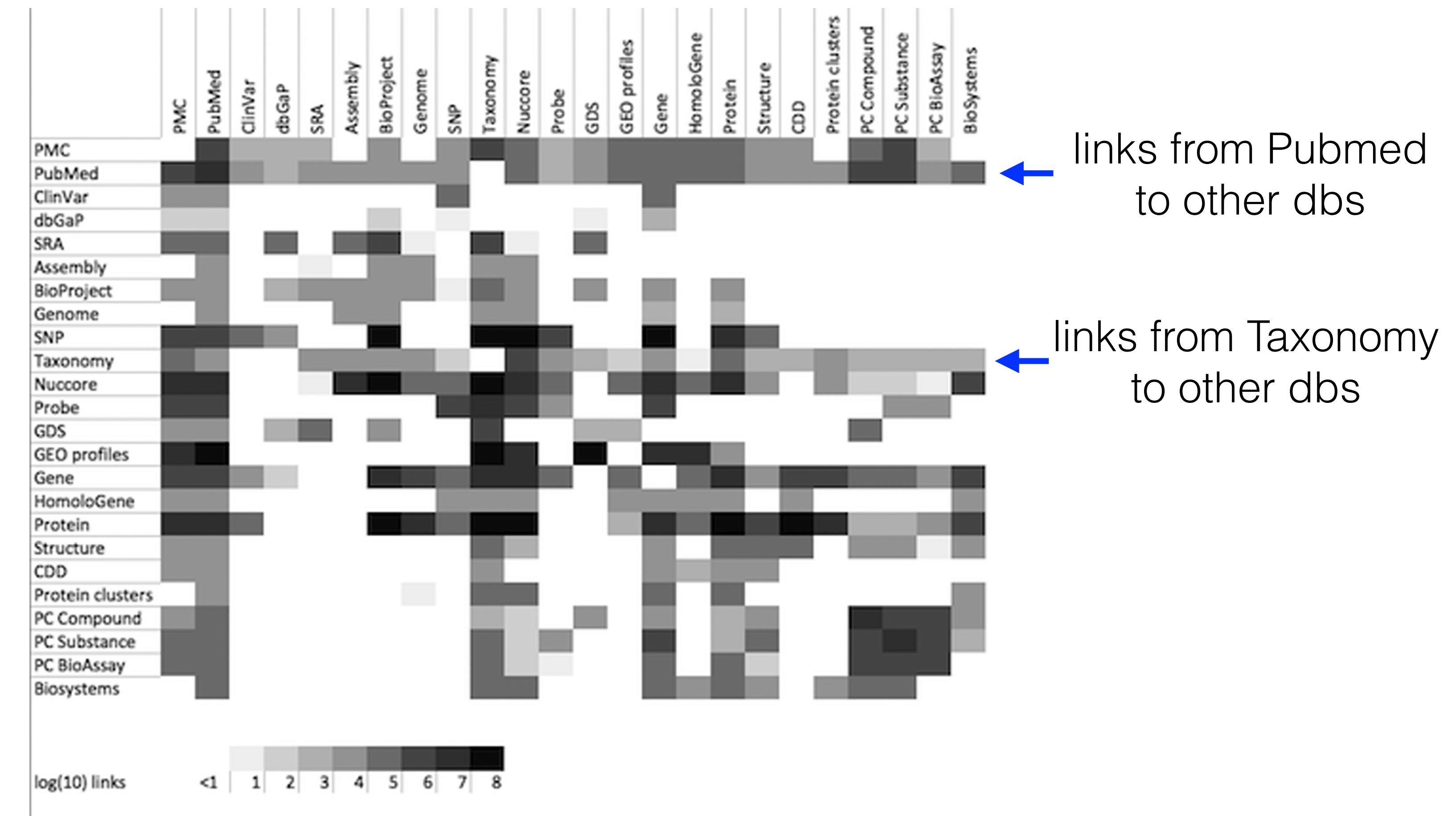
Today, we'll focus on NCBI databases

Category	Example NCBI db	Content
Literature	PubMed	Papers
Genome	Assembly	Genome assemblies
Taxonomy	Taxonomy	Information about species and higher order taxa
Proteins	Protein	Protein sequences
Plus many more...		

One useful feature of NCBI databases is that they connect to each other

So, you can, for example:

- get all the sequences associated with a species or genus
- get all the protein sequences encoded by a genome
- get the SRA datasets associated with a paper



The 2020 paper containing the original sequence description of SARS-CoV-2

Article

A new coronavirus associated with human respiratory disease in China

<https://doi.org/10.1038/s41586-020-2008-3>

Received: 7 January 2020

Accepted: 28 January 2020

Published online: 3 February 2020

Open access

 Check for updates

Fan Wu^{1,7}, Su Zhao^{2,7}, Bin Yu^{3,7}, Yan-Mei Chen^{1,7}, Wen Wang^{4,7}, Zhi-Gang Song^{1,7}, Yi Hu^{2,7}, Zhao-Wu Tao², Jun-Hua Tian³, Yuan-Yuan Pei¹, Ming-Li Yuan², Yu-Ling Zhang¹, Fa-Hui Dai¹, Yi Liu¹, Qi-Min Wang¹, Jiao-Jiao Zheng¹, Lin Xu¹, Edward C. Holmes^{1,5} & Yong-Zhen Zhang^{1,4,6}✉

Emerging infectious diseases, such as severe acute respiratory syndrome (SARS) and Zika virus disease, present a major threat to public health^{1–3}. Despite intense research efforts, how, when and where new diseases appear are still a source of considerable uncertainty. A severe respiratory disease was recently reported in Wuhan, Hubei province, China. As of 25 January 2020, at least 1,975 cases had been reported since the first patient was hospitalized on 12 December 2019. Epidemiological investigations have suggested that the outbreak was associated with a seafood market in Wuhan.

Here we study a single patient who was a worker at the market and who was admitted to the Central Hospital of Wuhan on 26 December 2019 while experiencing a severe respiratory syndrome that included fever, dizziness and a cough. Metagenomic RNA sequencing⁴ of a sample of bronchoalveolar lavage fluid from the patient identified a new RNA virus strain from the family *Coronaviridae*, which is designated here ‘WH-Human 1’ coronavirus (and has also been referred to as ‘2019-nCoV’).

Phylogenetic analysis of the complete viral genome (29,903 nucleotides) revealed that the virus was most closely related (89.1% nucleotide similarity) to a group of

The pubmed record for that paper

At the bottom of the pubmed page: related information links



Related information

[Assembly](#)

[Cited in Books](#)

[Domains](#)

[Gene](#)

[MedGen](#)

[Nucleotide](#) ←

[Nucleotide](#)

[Nucleotide \(Weighted\)](#)

[Protein](#)

[Protein \(RefSeq\)](#)

[Protein \(Weighted\)](#)

[Related Project](#)

[SRA](#)

[Taxonomy via GenBank](#)

Click this link to get to the actual virus genome sequence

We've jumped to the NCBI nucleotide database (Genbank)

The screenshot shows the NCBI Nucleotide database interface. At the top, there's a navigation bar with links for NCBI, Resources, How To, and Sign in to NCBI. Below the navigation is a search bar with dropdown menus for 'Nucleotide' and 'Advanced' search options, and a 'Search' button. A 'COVID-19 Information' sidebar is visible on the left.

The main content area displays the details for the 'Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome'. The NCBI Reference Sequence is listed as NC_045512.2. Below this, there are links for 'FASTA' and 'Graphics'.

A 'Go to:' dropdown menu is shown, containing the following items:

- LOCUS NC_045512
- DEFINITION Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome.
- ACCESSION NC_045512
- VERSION NC_045512.2
- DBLINK BioProject: [PRJNA485481](#)
- KEYWORDS RefSeq.
- SOURCE Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)
- ORGANISM Severe acute respiratory syndrome coronavirus 2
Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes;
Nidovirales; Cornidovirinae; Coronaviridae; Orthocoronavirinae;
Betacoronavirus; Sarbecovirus.
- REFERENCE 1 (bases 1 to 29903)
- AUTHORS Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y., Yuan, M.L., Zhang, Y.L., Dai, F.H., Liu, Y., Wang, Q.M., Zheng, J.J., Xu, L., Holmes, E.C. and Zhang, Y.Z.
- TITLE A new coronavirus associated with human respiratory disease in China
- JOURNAL [Nature](#) 570 (7762), 265-266 (2019)

To the right of the sequence details is a sidebar with various analysis tools and links:

- Send to: ▾ (Change region shown, Customize view)
- Analyze this sequence (Run BLAST, Pick Primers, Highlight Sequence Features, Find in this Sequence)
- NCBI Virus (Retrieve, view, and download SARS-CoV-2 coronavirus genomic and protein sequences)
- Related information (Assembly, BioProject, Protein, PubMed)

A blue arrow points to the 'Protein' link in the 'Related information' section.

Click this link to get to the protein sequences encoded by this genome

Exercise: download all the *protein* sequences encoded by this SARS-CoV-2 genome

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide Search Advanced Help

COVID-19 Information

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

GenBank Send to: ▾ Change region shown

Customize view

Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome

NCBI Reference Sequence: NC_045512.2

[FASTA](#) [Graphics](#)

Go to: ▾

LOCUS NC_045512 29903 bp ss-RNA linear VRL 18-JUL-2020

DEFINITION Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome.

ACCESSION NC_045512

VERSION NC_045512.2

DBLINK BioProject: [PRJNA485481](#)

KEYWORDS RefSeq.

SOURCE Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)

ORGANISM Severe acute respiratory syndrome coronavirus 2
Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes;
Nidovirales; Cornidovirinae; Coronaviridae; Orthocoronavirinae;
Betacoronavirus; Sarbecovirus.

REFERENCE 1 (bases 1 to 29903)

AUTHORS Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y.,
Tao, Z.W., Tian, J.H., Pei, Y.Y., Yuan, M.L., Zhang, Y.L., Dai, F.H.,
Liu, Y., Wang, Q.M., Zheng, J.J., Xu, L., Holmes, E.C. and Zhang, Y.Z.

TITLE A new coronavirus associated with human respiratory disease in China

JOURNAL [Nature](#) 570 (7762) 265-266 (2019)

Analyze this sequence Run BLAST Pick Primers Highlight Sequence Features Find in this Sequence

NCBI Virus Retrieve, view, and download SARS-CoV-2 coronavirus genomic and protein sequences.

Related information Assembly BioProject Protein PubMed

Click this link to get to the protein sequences encoded by this genome

Now we've jumped to the NCBI protein database

Sequence length
Custom range...

Molecular weight
Custom range...

Release date
Custom range...

Revision date
Custom range...

[Clear all](#)

[Show additional filters](#)

[ORF7b \[Severe acute respiratory syndrome coronavirus 2\]](#)
1. Accession: GI: 1820616061
[BioProject](#) [Nucleotide](#) [PubMed](#) [Taxonomy](#)
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)

[ORF1a polyprotein \[Severe acute respiratory syndrome coronavirus 2\]](#)
2. Accession: GI: 1802476803
[BioProject](#) [Nucleotide](#) [PubMed](#) [Taxonomy](#)
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)

[ORF10 protein \[Severe acute respiratory syndrome coronavirus 2\]](#)
3. 38 aa protein
Accession: YP_009725255.1 GI: 1798174256
[BioProject](#) [Nucleotide](#) [PubMed](#) [Taxonomy](#)
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)

[nucleocapsid phosphoprotein \[Severe acute respiratory syndrome coronavirus 2\]](#)
4. 419 aa protein
Accession: YP_009724397.2 GI: 1798174255
[BioProject](#) [Nucleotide](#) [PubMed](#) [Taxonomy](#)
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)

[ORF8 protein \[Severe acute respiratory syndrome coronavirus 2\]](#)
5. 121 aa protein
Accession: YP_009724396.1 GI: 1796318604
[BioProject](#) [Nucleotide](#) [PubMed](#) [Taxonomy](#)
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)

[ORF7a protein \[Severe acute respiratory syndrome coronavirus 2\]](#)
6. 121 aa protein
Accession: YP_009724395.1 GI: 1796318603
[BioProject](#) [Nucleotide](#) [PubMed](#) [Taxonomy](#)
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)

Find related data
Database: Select

[Find items](#)

Recent activity
[Turn Off](#) [Clear](#)

Protein Links for Nucleotide (Select 1798174254) (12) Protein

Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, c Nucleotide

Protein Links for Nucleotide (Select 1798172431) (10) Protein

Nucleotide (Weighted) Links for PubMed (Select 32015508) (411230) Nucleotide

Nucleotide Links for PubMed (Select 32015508) (2) Nucleotide

[See more...](#)

You could click on these sequences one at a time to access them

← → ⌂ https://www.ncbi.nlm.nih.gov/protein/YP_009724390.1

surface glycoprotein [Severe acute respiratory syndrome coronavirus 2]

NCBI Reference Sequence: YP_009724390

[Identical Proteins](#) [FASTA](#) [Graphics](#)

Go to:

LOCUS	YP_009724390	1273 aa	linear	VRL 18-JUL-2020
DEFINITION	surface glycoprotein [Severe acute respiratory syndrome coronavirus 2].			
ACCESSION	YP_009724390			
VERSION	YP_009724390.1			
DBLINK	BioProject: PRJNA485481			
DBSOURCE	REFSEQ: accession NC_045512.2			
KEYWORDS	RefSeq.			
SOURCE	Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)			
ORGANISM	Severe acute respiratory syndrome coronavirus 2 Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes; Nidovirales; Cornidovirineae; Coronaviridae; Orthocoronavirinae; Betacoronavirus; Sarbecovirus.			
REFERENCE	1 (residues 1 to 1273)			
AUTHORS	Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y., Yuan, M.L., Zhang, Y.L., Dai, F.H., Liu, Y., Wang, Q.M., Zheng, J.J., Xu, L., Holmes, E.C. and Zhang, Y.Z.			
TITLE	A new coronavirus associated with human respiratory disease in China			
JOURNAL	Nature 579 (7798), 265–269 (2020)			
PUBMED	32015508			
REMARK	Erratum: Nature 2020 Apr;580(7803):F7 PMID: 322061811			

Or you can download them all at once, in various formats

Summary ▾ 20 per page ▾ Sort by Default order ▾

Send to: ▾ Filters: [Manage Filters](#)

Items: 12

There were some problems retrieving the sequence. GI: 1820616061

There were some problems retrieving the sequence. GI: 1802476803

[ORF7b \[Severe acute respiratory syndrome coronavirus 2\]](#)

1. Accession: GI: 1820616061

[BioProject](#) [Nucleotide](#) [PubMed](#) [Taxonomy](#)

[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)

[ORF1a polyprotein \[Severe acute respiratory syndrome coronavirus 2\]](#)

2. Accession: GI: 1802476803

[BioProject](#) [Nucleotide](#) [PubMed](#) [Taxonomy](#)

[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)

[ORF10 protein \[Severe acute respiratory syndrome coronavirus 2\]](#)

3. 38 aa protein

Accession: YP_009725255.1 GI: 1798174256

[BioProject](#) [Nucleotide](#) [PubMed](#) [Taxonomy](#)

[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)

[nucleocapsid phosphoprotein \[Severe acute respiratory syndrome coronavirus 2\]](#)

4. 419 aa protein

Accession: YP_009724397.2 GI: 1798174255

[BioProject](#) [Nucleotide](#) [PubMed](#) [Taxonomy](#)

[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)

Choose Destination

- File Clipboard
 Collections Analysis Tool

Download 12 items.

Format

- Summary
- GenPept
- GenPept (full)
- FASTA
- ASN.1
- XML
- INSDSeq XML
- TinySeq XML
- Feature Table
- FASTA CDS
- Accession List
- GI List
- GFF3

sequences

with COBALT

ed Domains wi

nta

ct

Recent activity

Protein Links for Nucleotide 1798174254) (12)

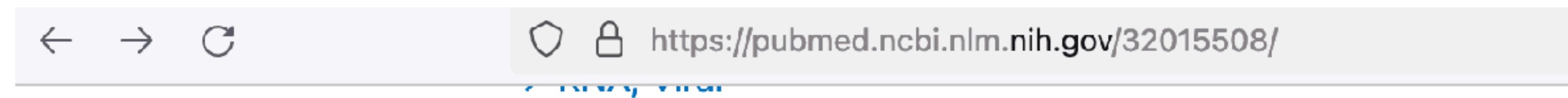
Severe acute respiratory sy coronaviru 2 isolate Wuha

Protein Links for Nucleotide 1798172431) (10)

Nucleotide (Weighted) Link (Select 32015508) (411230

Nucleotide Links for PubMed 32015508) (2)

At the bottom of the pubmed page: related information links



Related information

- [Assembly](#)
- [Cited in Books](#)
- [Domains](#)
- [Gene](#)
- [MedGen](#)
- [Nucleotide](#)
- [Nucleotide](#)
- [Nucleotide \(Weighted\)](#)
- [Protein](#)
- [Protein \(RefSeq\)](#)
- [Protein \(Weighted\)](#)
- [Related Project](#)
- [SRA](#)  Click this link to get to the NGS data from this paper
- [Taxonomy via GenBank](#)

The NCBI SRA database contains NGS datasets

NCBI Resources How To Sign in to NCBI

SRA SRA Search Advanced Help

COVID-19 Information

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

Full Send to:

Links from PubMed

SRX7636886: Complete genome of a novel coronavirus associated with severe human respiratory disease in Wuhan, China
1 ILLUMINA (Illumina MiniSeq) run: 28.3M spots, 8G bases, 2.6Gb downloads

Design: Total RNA was extracted from the BALF sample of a patient using the RNeasy Plus Universal Mini Kit (Qiagen) following the manufacturers instructions. An RNA library was then constructed using the SMARTer Stranded Total RNA-Seq Kit v2 (TaKaRa, Dalian, China). Ribosomal RNA (rRNA) depletion was performed during library construction following the manufacturers instructions. Paired-end (150 bp) sequencing of the RNA library was performed on the MiniSeq platform (Illumina).

Submitted by: Shanghai Public Health Clinical Center & School of Public Health, Fudan University

Study: Complete genome of a novel coronavirus associated with severe human respiratory disease in Wuhan, China
[PRJNA603194](#) • [SRP245409](#) • [All experiments](#) • [All runs](#)
[show Abstract](#)

Sample:
[SAMN13922059](#) • [SRS6067521](#) • [All experiments](#) • [All runs](#)
Organism: [human lung metagenome](#)

Library:
Name: 1
Instrument: Illumina MiniSeq
Strategy: RNA-Seq
Source: METATRANSCRIPTOMIC
Selection: RANDOM
Layout: PAIRED

Related Information

BioProject
BioSample
PMC
PubMed
Taxonomy

Recent activity

Turn Off Clear

SRA Links for PubMed (Select 32015508)
(1) SRA

Protein Links for Nucleotide (Select 1798174254) (12) Protein

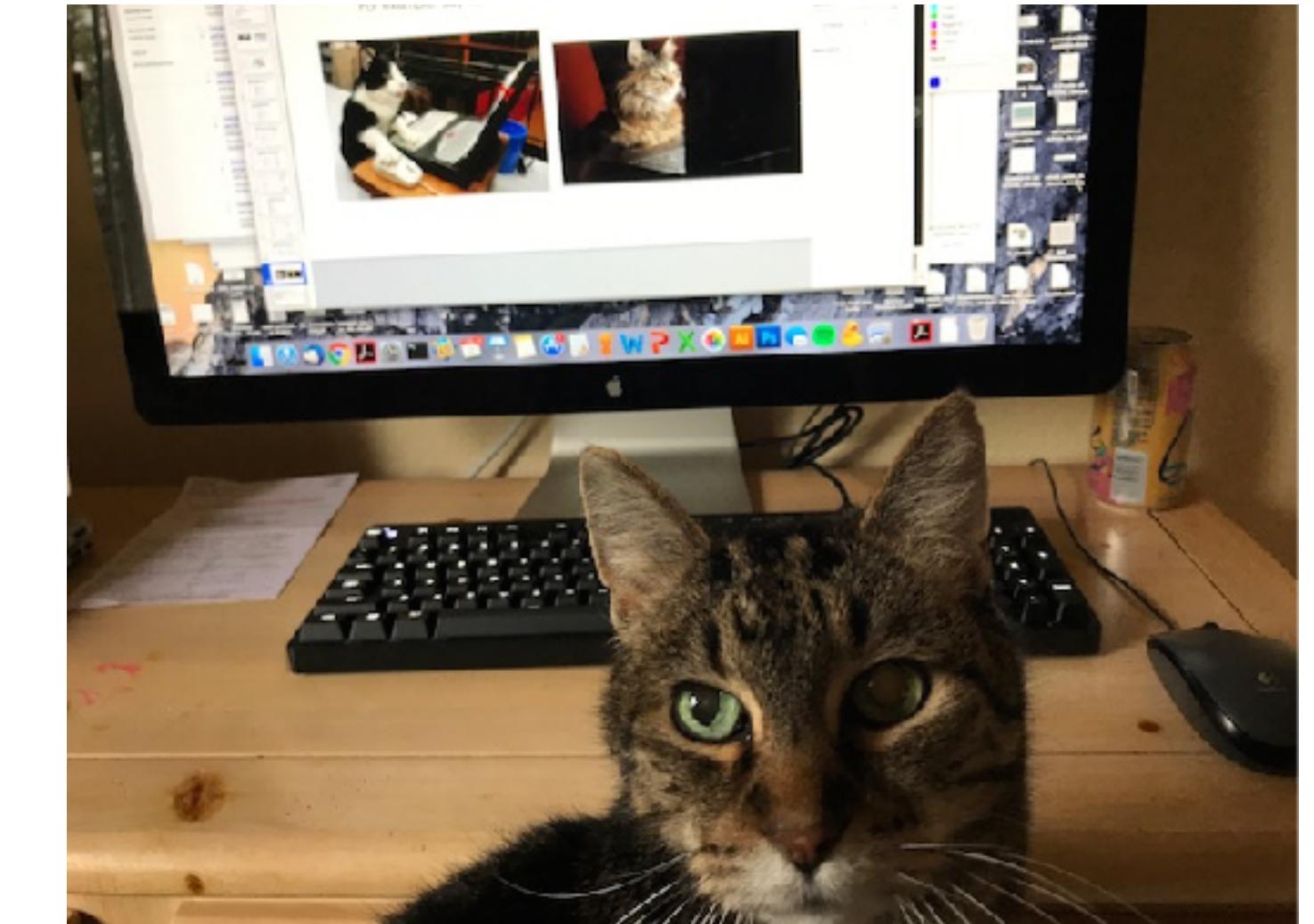
Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, c Nucleotide

Protein Links for Nucleotide (Select 1798172431) (10) Protein

Nucleotide (Weighted) Links for PubMed (Select 32015508) (411230) Nucleotide

A great way to find sequence data for an organism you are interested in is via the NCBI Taxonomy database

For example, say we want to download the cat (*Felis catus*) genome



Kirby

The Taxonomy database

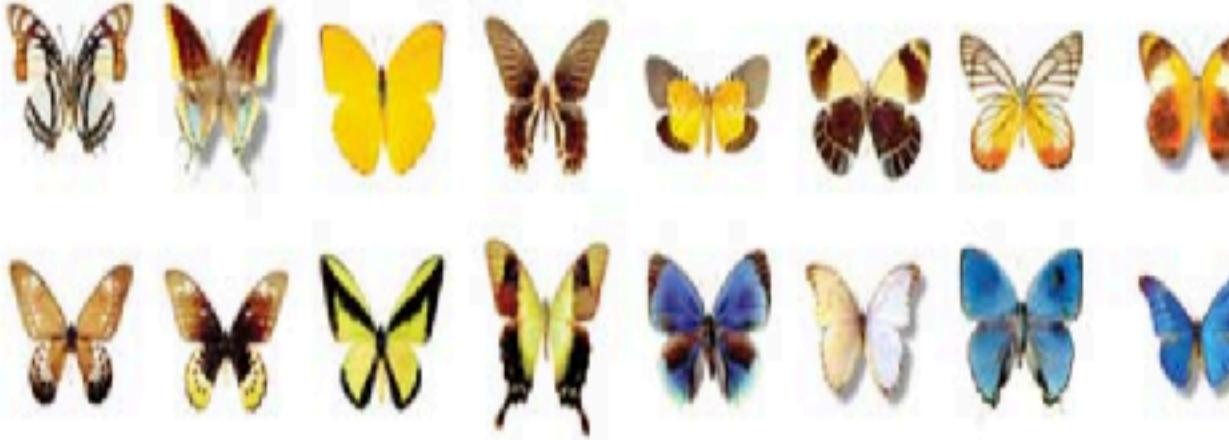
← → ⌂ https://www.ncbi.nlm.nih.gov/taxonomy ⌂ Sign in to NCBI

NCBI Resources How To

Taxonomy Taxonomy Search Limits Advanced Help

COVID-19 Information X

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



Taxonomy

The Taxonomy Database is a curated classification and nomenclature for all of the organisms in the public sequence databases. This currently represents about 10% of the described species of life on the planet.

Using Taxonomy

[Quick Start Guide](#)
[FAQ](#)
[Handbook](#)
[Taxonomy FTP](#)

Taxonomy Tools

[Browser](#)
[Common Tree](#)
[Statistics](#)
[Name/ID Status](#)
[Genetic Codes](#)
[Linking to Taxonomy](#)
[Extinct Organisms](#)

Other Resources

[GenBank](#)
[LinkOut](#)
[E-Utilities](#)
[Batch Entrez](#)
[INSDC](#)

The Taxonomy page for *Felis catus*

← → C https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=9685 ⭐ 📁 ↴ ↳ 🔍

NCBI  **Taxonomy Browser**

Entrez PubMed Nucleotide Protein Genome Structure PMC Taxonomy BioCollections

Search for as complete name lock

Display 3 levels using filter: none

Felis catus

Taxonomy ID: 9685 (for references in articles please use NCBI:txid9685)

— current name —

Felis catus Linnacus, 1758

| homotypic synonym: ***Felis silvestris catus***

| includes: **Korat cats** L.

Genbank common name: **domestic cat**

NCBI BLAST name: **carnivores**

Rank: **species**

Genetic code: [Translation table 1 \(Standard\)](#)

Mitochondrial genetic code: [Translation table 2 \(Vertebrate Mitochondrial\)](#)

Other names:

— heterotypic synonym —

Felis domesticus

— common name(s) —

cat, cats

Lineage ([full](#))

[cellular organisms](#); [Eukaryota](#); [Opisthokonta](#); [Metazoa](#); [Eumetazoa](#); [Bilateria](#); [Deuterostomia](#); [Chordata](#); [Craniata](#); [Vertebrata](#); [Gnathostomata](#); [Teleostomi](#); [Euteleostomi](#); [Sarcopterygii](#); [Dipnotetrapodomorpha](#); [Tetrapoda](#); [Amniota](#); [Mammalia](#); [Theria](#); [Eutheria](#); [Boreoeutheria](#); [Laurasiatheria](#); [Carnivora](#); [Feliformia](#); [Felidae](#); [Felinae](#); [Felis](#)

Entrez records

Database name	Direct links
Nucleotide	92,472
Protein	58,274
Structure	21
Genome	1
Popset	207
GEO Datasets	277
PubMed Central	3,386
Gene	46,051
SRA Experiments	2,492
Protein Clusters	12
Identical Protein Groups	45,451
Bio Project	110
Bio Sample	1,649
Bio Systems	495
Assembly	8
Probe	2,877
PubChem BioAssay	1,118
Taxonomy	1

genome ← SRA datasets ←

Felis catus in the NCBI genome database

https://www.ncbi.nlm.nih.gov/genome/?term=txid9685[Organism:noexp]

Genome Genome txid9685[Organism:noexp] Search Create alert Limits Advanced Help

COVID-19 Information X

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

Felis catus (domestic cat)
Reference genome: [Felis catus \(assembly Felis_catus_9.0\)](#)
Download sequences in FASTA format for [genome](#), [transcript](#), [protein](#) ←
Download genome annotation in [GFF](#), [GenBank](#) or [tabular](#) format
BLAST against Felis catus [genome](#), [transcript](#), [protein](#)

All 4 genomes for species:
[Browse the list](#)
[Download sequence and annotation from RefSeq or GenBank](#)
NEW Try [NCBI Datasets](#) - a new way to download genome sequence and annotation we're testing in NCBI Labs

Display Settings: Overview Send to: ID: 78

Organism Overview ; [Genome Assembly and Annotation report \[4\]](#) ; [Organelle Annotation Report \[1\]](#)

Felis catus (domestic cat)
domestic cat

 Lineage: [Eukaryota](#)[7836]; [Metazoa](#)[3708]; [Chordata](#)[1775]; [Craniata](#)[1753]; [Vertebrata](#)[1753]; [Euteleostomi](#)[1737]; [Mammalia](#)[470]; [Eutheria](#)[445]; [Laurasiatheria](#)[255]; [Carnivora](#)[65]; [Feliformia](#)[24]; [Felidae](#)[16]; [Felinae](#)[11]; [Felis](#)[2]; [Felis catus](#)[1]

Felis catus, the domestic cat, provides several valuable models for infectious disease, including a model for human AIDS. With a large number of recognized breeds, the cat is also a valuable resource for studying phenotypic diversity and evolution. The cat genome will further facilitate research in human medicine as some rare diseases that occur [More...](#)

Summary

Sequence data: genome assemblies: 4; sequence reads: 2 (See [Genome Assembly and Annotation report](#))
Statistics: median total length (Mb): 2507.5
median protein count: 54726
median GC%: 41.8903

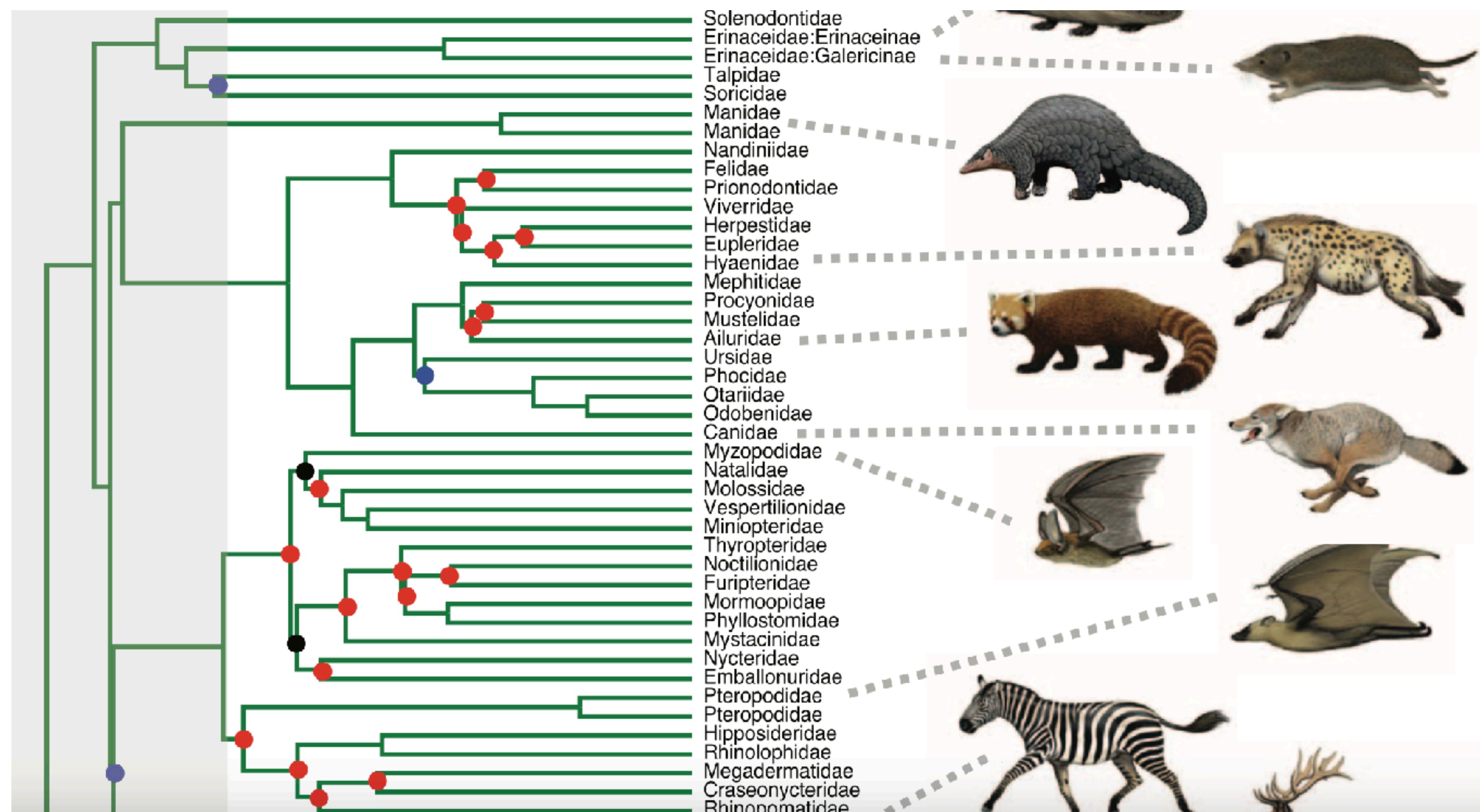
NCBI Resources
[Genome Data Viewer](#)

Tools
[BLAST Genome](#)

Related Information
[Assembly](#)
[BioProject](#)
[Gene](#)
[Components](#)
[Protein](#)
[PubMed](#)
[Taxonomy](#)

Search details
txid9685 [Organism:noexp]

The taxonomy database includes the taxonomic lineage of organisms



You can go to any point in the taxonomic tree of life in the Taxonomy db

Felis catus

Taxonomy ID: 9685 (for references in articles please use NCBI:txid9685)

current name

Felis catus Linnaeus, 1758

homotypic synonym: ***Felis silvestris catus***

includes: ***Korat cats*** L.

Genbank common name: **domestic cat**

NCBI BLAST name: **carnivores**

Rank: **species**

Genetic code: [Translation table 1 \(Standard\)](#)

Mitochondrial genetic code: [Translation table 2 \(Vertebrate Mitochondrial\)](#)

Other names:

heterotypic synonym

Felis domesticus

common name(s)

cat, cats

[Lineage](#)(full)

[cellular organisms](#); [Eukaryota](#); [Opisthokonta](#); [Metazoa](#); [Eumetazoa](#); [Bilateria](#); [Deuterostomia](#); [Chordata](#); [Craniata](#); [Vertebrata](#); [Gnathostomata](#); [Teleostomi](#); [Euteleostomi](#); [Sarcopterygii](#); [Dipnotetrapodomorpha](#); [Tetrapoda](#); [Amniota](#); [Mammalia](#); [Theria](#); [Eutheria](#); [Boreoeutheria](#); [Laurasiatheria](#); [Carnivora](#); [Feliformia](#); [Felidae](#); [Felinae](#); [Felis](#)

Eukaryotes

Mammals

Felidae

All the available genomes for species in the Felidae family

Search for as complete name lock

Display 3 levels using filter: none

Nucleotide Protein Structure Genome Popset SNP Conserved Domains GEO Datasets PubMed Central
 Gene HomoloGene SRA Experiments LinkOut BLAST GEO Profiles Protein Clusters Identical Protein Groups SPARCLE
 Bio Project Bio Sample Bio Systems Assembly dbVar Genetic Testing Registry Host Viral Host Probe
 PubChem BioAssay

Lineage (full): cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Dipnotetrapodomorpha; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Boreoeutheria; Laurasiatheria; Carnivora; Feliformia

- **Felidae** (cat family) [38](#) Click organism name to get more information 38 genomes for species in Felidae
 - [Acinonychinae](#) [1](#)
 - [Acinonyx](#) [1](#)
 - [Acinonyx jubatus](#) (cheetah) [1](#)
 - [Felinae](#) [30](#)
 - [Caracal](#) [1](#)
 - [Caracal caracal](#) [1](#)
 - [Catopuma](#) [2](#)
 - [Catopuma badia](#) (bay cat) [1](#)
 - [Catopuma temminckii](#) (Asiatic golden cat) [1](#)
 - [Felinae intergeneric hybrids](#) [1](#)
 - [Felis catus x Leopardis geoffroyi](#) [1](#)
 - [Felis catus x Prionailurus bengalensis](#)
 - [Leptailurus serval x Caracal caracal](#)
 - [Felis](#) [5](#)
 - [Felis catus](#) (domestic cat) [1](#)
 - [Felis chaus](#) (jungle cat) [1](#)
 - [Felis chaus x Felis catus](#)
 - [Felis margarita](#) (sand cat) [1](#)
 - [Felis nigripes](#) (black-footed cat) [1](#)
 - [Felis silvestris](#) (wild cat) [1](#)
 - [unclassified Felis](#)
 - [environmental samples](#)
 - [Leopardus](#) [7](#)

Felidae genomes

https://www.ncbi.nlm.nih.gov/genome/?term=txid9681[Organism:exp]

NCBI Resources How To Sign in to NCBI

Genome Genome txid9681[Organism:exp] Search Create alert Limits Advanced Help

COVID-19 Information X

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

See also 22 organelle- and plasmid-only records matching your search

Display Settings: Summary, 20 per page Send to: Database: Select Find items

Search results

Items: 16

[Panthera tigris](#)
1. tiger
Kingdom: Eukaryota; Subgroup: Mammals
Sequence data: genome assemblies:3
Haploid chromosomes: 19; Organelles: 1
Date: 2013/09/05
ID: 10802

[Panthera leo](#)
2. [Panthera leo overview](#)
Kingdom: Eukaryota; Subgroup: Mammals
Sequence data: genome assemblies:3
Haploid chromosomes: 19
Date: 2019/10/01
ID: 13342

Filters: [Manage Filters](#)

Find related data

Search details

txid9681[Organism:exp]

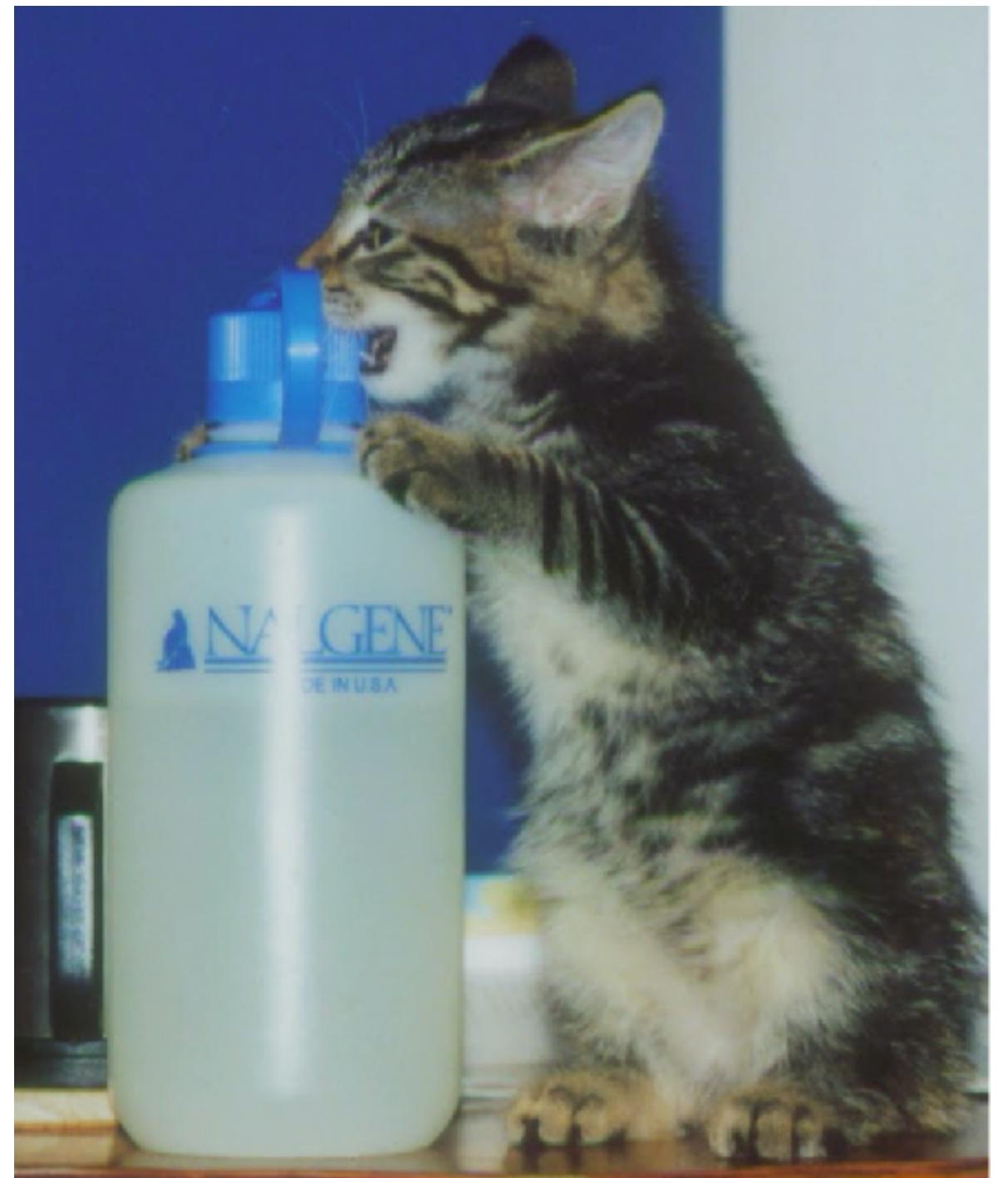
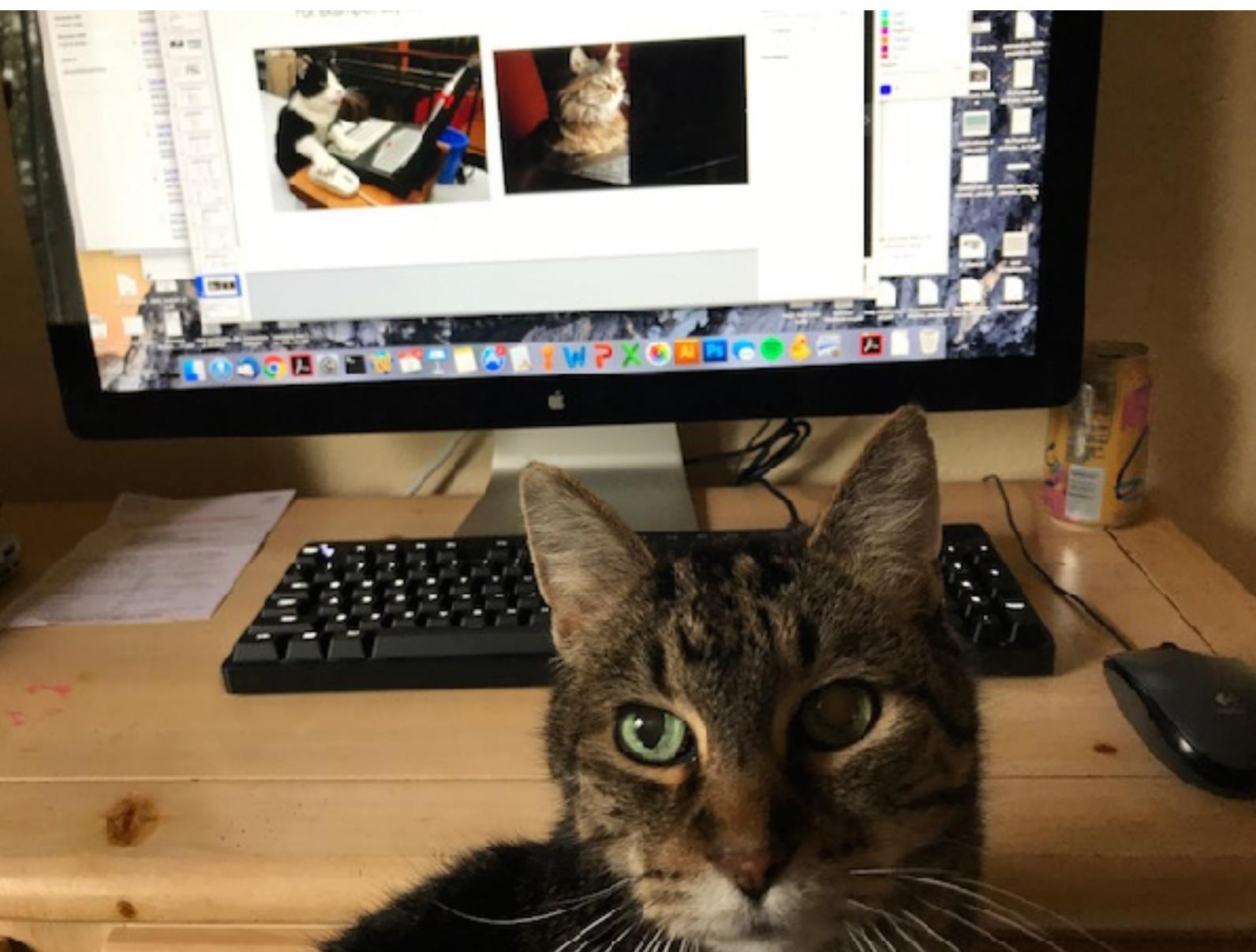
Recent activity

Turn Off Clear

[txid9681\[Organism:exp\] \(16\)](#) Genome

[felis catus \(1\)](#) Taxonomy

Questions?



Kirby in 2000