

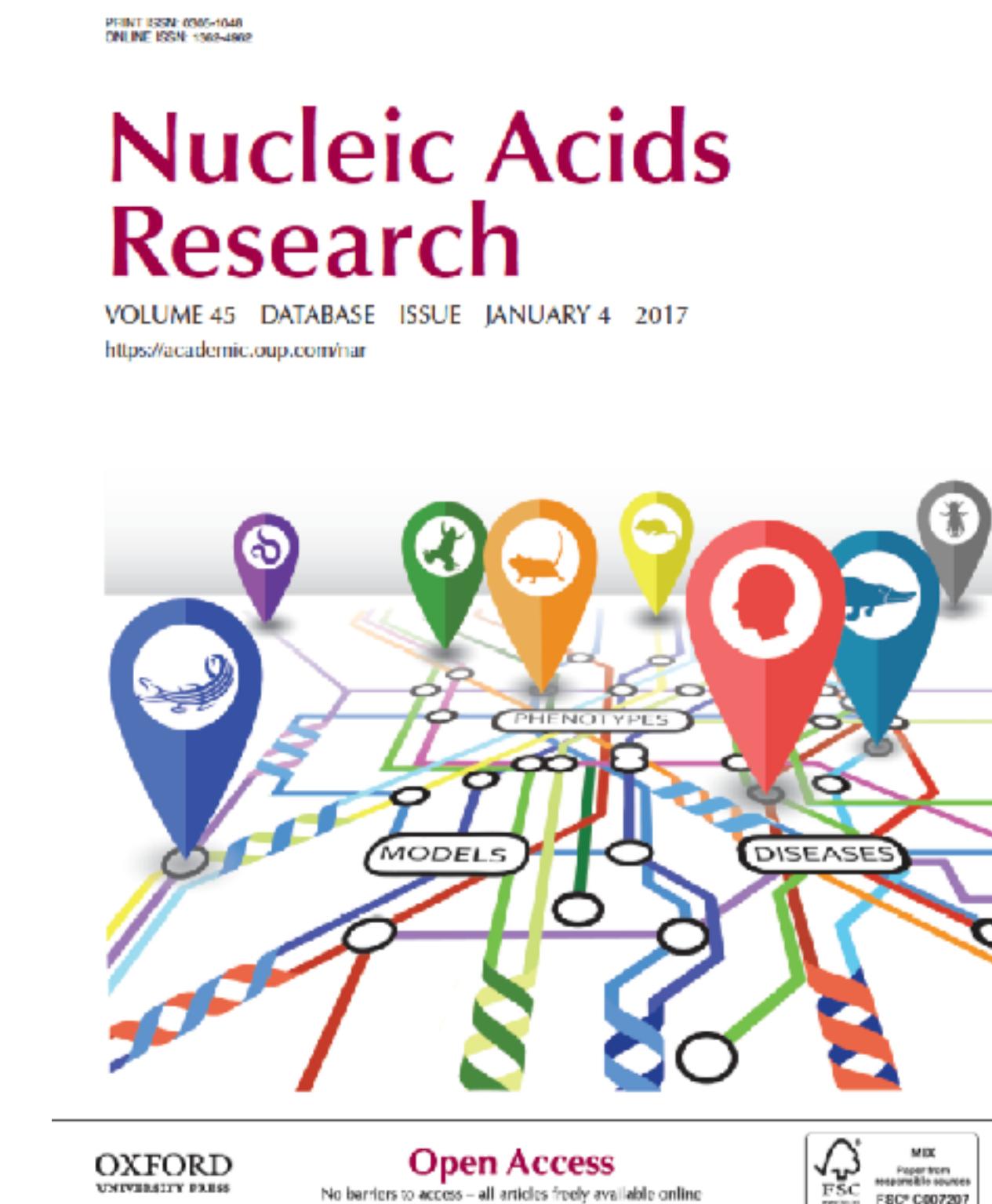
An overview of bioinformatics databases and online resources: what they are and how to access them

Mark Stenglein, GDW 2018



There are an overwhelming number of databases and other online resources, which often have overlapping content and purpose

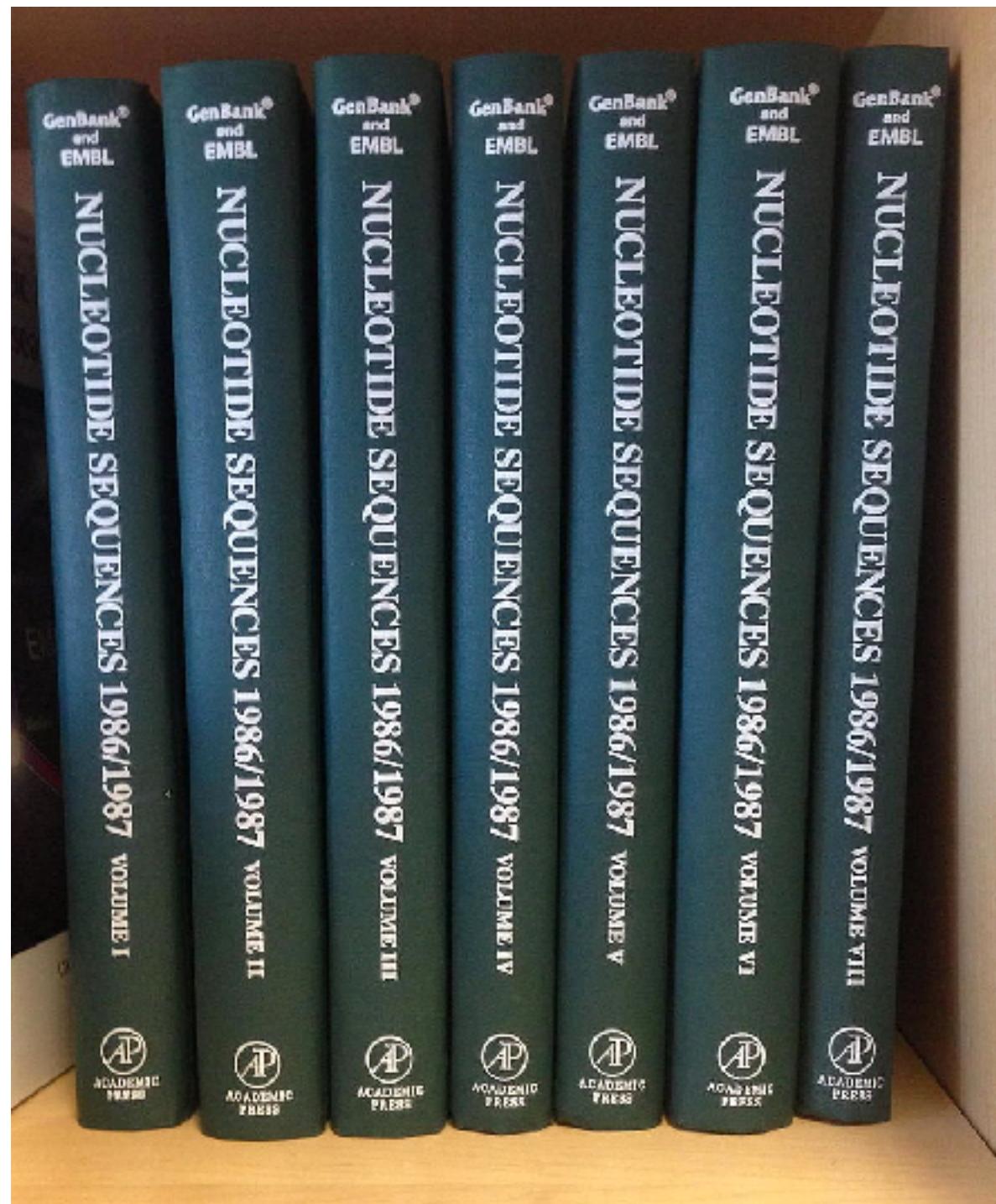
The annual Database and Web Server NAR issue is a good resource



<https://academic.oup.com/nar/issue/45/D1>

GenBank was one of the earliest sequence databases.

GenBank circa 1987



~10,000 sequences

GenBank release 100 (1997)

distributed by CDROM



~1,300,000 sequences

Genbank today



>200,000,000 sequences

BOVCHYMOA						NUCLEOTIDE SEQUENCES 1984					
SITES:						key			description		
key	site	span	description			key	site	span	description		
refnumbr	21	1	numbered 1 in [1]			pept/pept	195	0	chymo _a propept end/ mature pept		
->pept	21	1	chymo _a prepropept cds start						start		
pept/pept	69	0	chymo _a prepropept end/ propept start			pept<-	1166	1	chymo _a mature pept cds end		
ORIGIN:	20 bases upstream from codon 1										
SEQUENCE:	1275	bp	293	a	391	c	336	g	255	t	
1	ccgttggatc	cagatccaa	atggagggtgc	tgcgttgtgt	acttgttgtc	ttcgtctttt	ccc	aaaggccgc	ttagatccac	aggatccctc	tgtacaaaa
101	caagtcttcg	aggaaaggcc	tgcgggggtc	tgggttttcgg	ggggacttcc	tgcggaaaata	ccgtatggc	atccggcgca	atgtactcccg	tttcgggggg	aaag
201	gtggccacgg	tgc	ccccctgtac	caactacactg	gataggctgt	actttggggaa	gtttttatcc	ggggatcccg	ttccagggtt	cccggtctgt	tttgacactg
301	gttccctctga	cttctgggtt	cccttatact	atgcggatgg	caatggcttc	aaaaaccacc	ggcggttcga	cccgagaaag	tgttccaccc	tccagaaatc	
401	ggggcaagccc	ctgttatacc	actacggggc	aggcgacatg	caggggatcc	tgggttatgt	cccggttact	gttcccccac	tttgcgacat	ccagcaga	
501	gttgggttgc	gcacccatgg	gccccggggc	gtttccaccc	atgcggaaat	cgatggggatc	ctggggatgg	ccatccccc	gttgcgttcc	gatgtact	
601	tacccgttgt	tgacacatgt	atggacagge	atcggttgc	ccaagaccc	tttgcgtttt	acatggacac	gtatggccag	ggggatgttc	tcacgttgg	
701	ggccatcgac	ctgtgtact	acacagggtc	ccgttactgg	gttgcgttgc	ctgtgtgtca	gtatggccag	tttgcgtttt	ggggatgttc	tcacgttgg	
801	gttgggttgtt	cttgcgtggg	tgggtgttc	ggatgttc	gttgcgttgc	cttccaaatgt	gtatggccca	ccatccatcc	ggggatgttc	tcacgttgg	
901	ttggggccac	acggaaacc	tgatgtgt	tttgcgttgc	tttgcgttgc	cttgcgttca	tgccatgt	gttgcgttgc	tttgcgttgc	aaatgtaccc	
1001	actgacccccc	tttgcgttata	ccatccatgg	ccatgggttc	tgttccatgt	gttgcgttgc	atccatgt	tttgcgttgc	tttgcgttgc	tttgcgttgc	
1101	atcgatggat	tttgcgttgc	ccatccatgg	ccatgggttc	tgttccatgt	gttgcgttgc	atccatgt	tttgcgttgc	tttgcgttgc	tttgcgttgc	
1201	acatccatgt	acatccatgt	acatccatgt	atggcacatgt	tgttccatgt	gttgcgttgc	atccatgt	tttgcgttgc	tttgcgttgc	tttgcgttgc	

Today, we'll focus mainly on NCBI databases and resources, and how to access them

The NCBI was created in 1987 by the US government

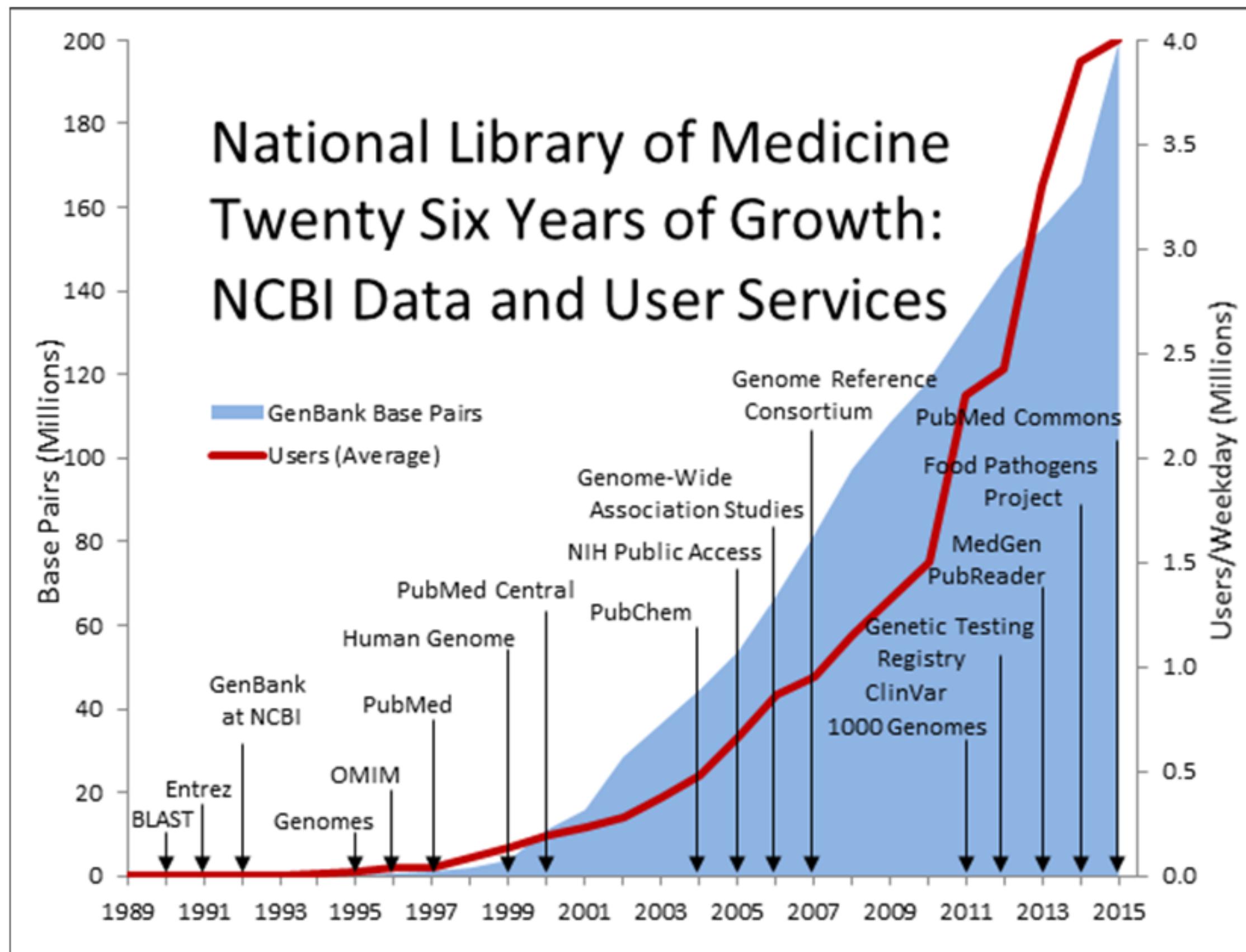


image: NIH/NLM

Categories of NCBI databases

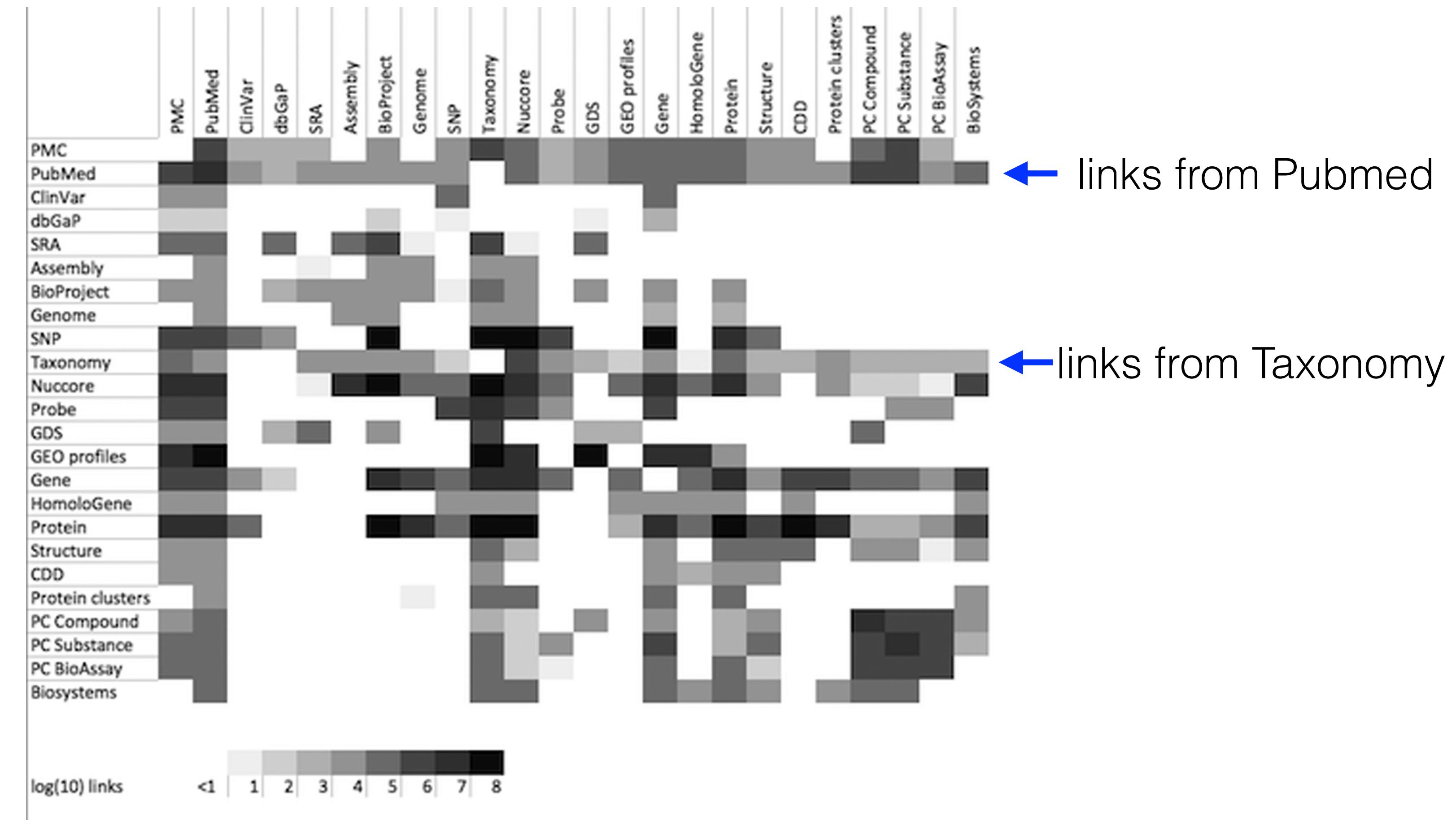
Category	Example NCBI db	Content
Literature	PubMed	Scientific and medical abstracts/citations
Genomes	Assembly	Genome assembly information
Genes	Gene	Collected information about gene loci
Proteins	Protein	Protein sequences
Chemicals	PubChem Compound	Chemical information with structures, information and links
Health	dbGaP	Genotype/phenotype interaction studies

<https://academic.oup.com/nar/issue/45/D1>

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

So, you can, for example:

- get all the nucleotide sequences associated with a taxon of interested
- get all the protein sequences predicted to be encoded by a genome
- get the SRA datasets associated with a particular paper in Pubmed



Say we want to find Bob's papers on panthers and camels

You can search NCBI databases using boolean operators: AND, OR, NOT

The screenshot shows the PubMed search results page. The search query is "Fitak AND panther or camel". The results are displayed in a table with columns for Article types, Format, Sort by, Per page, and Send to. The search results section shows "Items: 1 to 20 of 4703". A blue arrow points to this text. Below it, there are three search results listed:

Article types	Format: Summary	Sort by: Most Recent	Per page: 20	Send to
Clinical Trial				
Review				
Customize ...				
Text availability	Items: 1 to 20 of 4703	<< First < Prev Page 1 of 236 Next > Last >>		
Abstract				
Free full text				
Full text				
PubMed Commons	<input type="checkbox"/> A fossil protein chimera; difficulties in discriminating dinosaur peptide sequences from modern cross-contamination.			
Reader comments	1. Buckley M, Warwood S, van Dongen B, Kitchener AC, Manning PL. Proc Biol Sci. 2017 May 31;284(1855). pii: 20170544. doi: 10.1098/rspb.2017.0544.			
Trending articles	PMID: 28566488			
Publication dates	<input type="checkbox"/> Effects of dexamethasone on progesterone and estrogen profiles and uterine progesterone receptor localization during pregnancy in Sahel goat in Semi-Arid region.			
5 years	2. Yahi D, Ojo NA, Mshelia GD. J Anim Sci Technol. 2017 May 26;59:12. doi: 10.1186/s40781-017-0137-5. eCollection 2017.			
10 years	PMID: 28560047	Free Article		
Custom range...	Similar articles			
Species	<input type="checkbox"/> Physico-chemical, microstructural and rheological properties of camel-milk yogurt as enhanced by microbial transglutaminase.			
Humans	3. Abou-Soliman NHI, Sakr SS, Awad S. J Food Sci Technol. 2017 May;54(6):1616-1627. doi: 10.1007/s13197-017-2593-9. Epub 2017 Mar 27.			
Other Animals	PMID: 28559621	Similar articles		
Clear all				
Show additional filters				

Use parentheses to group search terms

NCBI Resources ▾ How To ▾

PubMed.gov US National Library of Medicine National Institutes of Health

PubMed Fitak AND (panther or camel) ←

Create RSS Create alert Advanced

Article types Clinical Trial Review Customize ...

Text availability Abstract Free full text Full text

PubMed Commons Reader comments Trending articles

Publication dates 5 years 10 years Custom range...

Species Humans Other Animals

[Clear all](#)

[Show additional filters](#)

Format: Summary ▾ Sort by: Most Recent ▾ Send to ▾

Search results Items: 4 ←

[Evolutionary and Functional Mitogenomics Associated With the Genetic Restoration of the Florida Panther.](#)
1. [Panther.](#)
Ochoa A, Onorato DP, **Fitak RR**, Roelke-Parker ME, Culver M.
J Hered. 2017 Jun 1;108(4):449-455. doi: 10.1093/jhered/esx015.
PMID: 28204600 [Similar articles](#)

[The major histocompatibility complex in Old World camelids and low polymorphism of its class II genes.](#)
2. [genes.](#)
Plasil M, Mohandesan E, **Fitak RR**, Musilova P, Kubickova S, Burger PA, Horin P.
BMC Genomics. 2016 Mar 1;17:167. doi: 10.1186/s12864-016-2500-1.
PMID: 26931144 [Free PMC Article](#) [Similar articles](#)

[The de novo genome assembly and annotation of a female domestic dromedary of North African origin.](#)
3. [origin.](#)
Fitak RR, Mohandesan E, Corander J, Burger PA.
Mol Ecol Resour. 2016 Jan;16(1):314-24. doi: 10.1111/1755-0998.12443. Epub 2015 Jul 24.
PMID: 26178449 [Free PMC Article](#) [Similar articles](#)

[Diagnostic single nucleotide polymorphism markers to identify hybridization between dromedary and Bactrian camels.](#)
4. [and Bactrian camels.](#)
Ruiz E, Mohandesan E, **Fitak RR**, Burger PA.
Conserv Genet Resour. 2015;7(2):329-332. Epub 2015 Jan 13.
PMID: 26161164 [Free PMC Article](#)

Get nucleotide sequences associated with these papers

[https://www.ncbi.nlm.nih.gov/pubmed/?term=Fitak+AND+\(panther+or+camel\)](https://www.ncbi.nlm.nih.gov/pubmed/?term=Fitak+AND+(panther+or+camel))

NCBI Resources ▾ How To ▾ Sign in to NCBI

PubMed Fitak AND (panther or camel) Search Create RSS Create alert Advanced Help

Article types Clinical Trial Review Customize ...

Text availability Abstract Free full text Full text

PubMed Commons Reader comments Trending articles

Publication dates 5 years 10 years Custom range...

Species Humans Other Animals

[Clear all](#)

[Show additional filters](#)

Format: Summary ▾ Sort by: Most Recent ▾ Send to ▾ Filters: [Manage Filters](#)

Find related data Database: Select

Find items

Search details

Fitak[All Fields] AND (("panthera" [MeSH Terms] OR "panthera" [All Fields] OR "panther" [All Fields]) OR ("camelus" [MeSH Terms] OR

Recent Activity Turn Off Clear

Fitak AND (panther OR camel) (4) PubMed

Fitak AND panther OR camel (4703) PubMed

Fitak RR (10) PubMed

Fitak (33) PubMed

Fitak AND (dromedary) (3) PubMed

See more...

Search See more...

Search results Items: 4

1. [Evolutionary and Functional Mitogenomics Associated With the Genetic Restoration of the Florida Panther.](#)
Ochoa A, Onorato DP, **Fitak RR**, Roelke-Parker ME, Culver M.
J Hered. 2017 Jun 1;108(4):449-455. doi: 10.1093/jhered/esx015.
PMID: 28204600
[Similar articles](#)

2. [The major histocompatibility complex in Old World camelids and low polymorphism of its class II genes.](#)
Plasil M, Mohandesan E, **Fitak RR**, Musilova P, Kubickova S, Burger PA, Horin P.
BMC Genomics. 2016 Mar 1;17:167. doi: 10.1186/s12864-016-2500-1.
PMID: 26931144 [Free PMC Article](#)
[Similar articles](#)

3. [The de novo genome assembly and annotation of a female domestic dromedary of North African origin.](#)
Fitak RR, Mohandesan E, Corander J, Burger PA.
Mol Ecol Resour. 2016 Jan;16(1):314-24. doi: 10.1111/1755-0998.12443. Epub 2015 Jul 24.
PMID: 26178449 [Free PMC Article](#)
[Similar articles](#)

4. [Diagnostic single nucleotide polymorphism markers to identify hybridization between dromedary and Bactrian camels.](#)
Ruiz E, Mohandesan E, **Fitak RR**, Burger PA.
Conserv Genet Resour. 2015;7(2):329-332. Epub 2015 Jan 13.
PMID: 26161164 [Free PMC Article](#)
[Similar articles](#)

Find related data Database: Select ←

Get nucleotide sequences associated with these publications

[https://www.ncbi.nlm.nih.gov/pubmed/?term=Fitak+AND+\(panther+or+camel\)](https://www.ncbi.nlm.nih.gov/pubmed/?term=Fitak+AND+(panther+or+camel))

Summary ▾ 20 per page ▾ Sort by Default order ▾

Send: ▾

Filters: [Manage Filters](#)

Items: 1 to 20 of 26

<< First < Prev Page 1 of 2 Next > Last >>

- [Camelus ferus MHC class II antigen \(Cafe-DQA\) gene, Cafe-DQA*02 allele, exon 2 and partial cds](#)

1. 249 bp linear DNA

Accession: KT936421.1 GI: 1003311003

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

- [Camelus ferus MHC class II antigen \(Cafe-DQA\) gene, Cafe-DQA*01 allele, exon 2 and partial cds](#)

2. 249 bp linear DNA

Accession: KT936420.1 GI: 1003311001

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

- [Camelus dromedarius MHC class II antigen \(Cadr-DQA\) gene, Cadr-DQA*01 allele, exon 2 and partial cds](#)

3. 249 bp linear DNA

Accession: KT936419.1 GI: 1003310999

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

- [Camelus bactrianus MHC class II antigen \(Caba-DQA\) gene, Caba-DQA*03 allele, exon 2 and partial cds](#)

4. 249 bp linear DNA

Accession: KT936418.1 GI: 1003310997

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

Results by taxon

Top Organisms [Tree]

[Camelus bactrianus](#) (10)
[Camelus dromedarius](#) (9)
[Camelus ferus](#) (7)

Find related data

Database: [Select](#)

[Find items](#)

Recent activity

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

Diagnostic single nucleotide polymorphism markers to identify hybridization betw.. [PubMed](#)

(Fitak and AND (camelid OR dromedary)) OR (Stenglein AND (snake O... (5) [PubMed](#)

(Fitak and AND (camelid OR dromedary)) OR (Stenglein AND (snake O... (13) [PubMed](#)

Fitak and AND (camelid OR dromedary) (3) [PubMed](#)

You could click on these sequences one at a time

Secure | <https://www.ncbi.nlm.nih.gov/nuccore/KT936421.1>

NCBI Resources ▾ How To ▾

Nucleotide Nucleotide ▾ Advanced

GenBank ▾ Send: ▾

Camelus ferus MHC class II antigen (Cafe-DQA) gene, Cafe-DQA*02 allele, exon 2 and partial cds

GenBank: KT936421.1

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

Go to: ▾

LOCUS KT936421 249 bp DNA linear MAM 13-MAR-2016

DEFINITION Camelus ferus MHC class II antigen (Cafe-DQA) gene, Cafe-DQA*02 allele, exon 2 and partial cds.

ACCESSION KT936421

VERSION KT936421.1

KEYWORDS .

SOURCE Camelus ferus (Wild Bactrian camel)

ORGANISM [Camelus ferus](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Tylopoda;
Camelidae; Camelus.

REFERENCE 1 (bases 1 to 249)

AUTHORS Plasil,M., Mohandesan,E., Fitak,R.R., Musilova,P., Kubickova,S.,
Burger,P.A. and Horin,P.

TITLE The major histocompatibility complex in Old World camelids and low
polymorphism of its class II genes

JOURNAL BMC Genomics 17 (1), 167 (2016)

PUBMED [26931144](#)

REMARK Publication Status: Online-Only

REFERENCE 2 (bases 1 to 249)

AUTHORS Plasil,M.

TITLE Direct Submission

JOURNAL Submitted (16-OCT-2015) Animal Immunogenomics, CEITEC, VFU,
Palackeho trida 1/3, Brno 61242, Czech Republic

COMMENT ##Assembly-Data-START##
Assembly Method :: BWA v. 0.6.2
Sequencing Technology :: 454; Illumina; Sanger dideoxy sequencing
##Assembly-Data-END##

FEATURES Location/Qualifiers

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

Nucleotide Help

Summary ▾ 20 per page ▾ Sort by Default order ▾  Send: ▾ Filters: [Manage Filters](#)

Items: 1 to 20 of 26

<< First < Prev Page 1 of 2 Next > Last >>

- [Camelus ferus MHC class II antigen \(Cafe-DQA\) gene, Cafe-DQA*02 allele, exon 2 and partial cds](#)
 - 1. 249 bp linear DNA
Accession: KT936421.1 GI: 1003311003
[GenBank](#) [FASTA](#) [Graphics](#)
 - 2. 249 bp linear DNA
Accession: KT936420.1 GI: 1003311001
[GenBank](#) [FASTA](#) [Graphics](#)
 - 3. 249 bp linear DNA
Accession: KT936419.1 GI: 1003310999
[GenBank](#) [FASTA](#) [Graphics](#)
 - 4. 249 bp linear DNA
Accession: KT936418.1 GI: 1003310997
[GenBank](#) [FASTA](#) [Graphics](#)

Results by taxon 

Top Organisms [Tree]

- [Camelus bactrianus \(10\)](#)
- [Camelus dromedarius \(9\)](#)
- [Camelus ferus \(7\)](#)

Find related data 

Database:

Recent activity 

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

-  Diagnostic single nucleotide polymorphism markers to identify hybridization betw.. PubMed
-  (Fitak and AND (camelid OR dromedary)) OR (Stenglein AND (snake O... (5) PubMed
-  (Fitak and AND (camelid OR dromedary)) OR (Stenglein AND (snake O... (13) PubMed
-  Fitak and AND (camelid OR dromedary) (3) PubMed

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

https://www.ncbi.nlm.nih.gov/nuccore?linkname=pubmed_nuccore&querykey=14

The screenshot shows the NCBI Nucleotide search results page for species Camelus. The results list four entries, each with a checkbox and a link to the sequence details. A dropdown menu is open on the right side of the page, titled 'Send:'. It contains several options under 'Format':

- Complete Record (radio button selected)
- Coding Sequences
- Gene Features
- Choose Destination:
 - File (radio button selected)
 - Clipboard
 - Collections
- Download 26 items.
- Format:
 - ✓ Summary
 - GenBank
 - GenBank (full)
 - FASTA (selected)
 - ASN.1
 - XML
 - INSDSeq XML
 - TinySeq XML
 - Feature Table
 - Accession List
 - GI List
 - GFF3

Two blue arrows point from the text 'includes annotations' and 'just the sequences' to the 'Complete Record' and 'FASTA' options respectively in the dropdown menu.

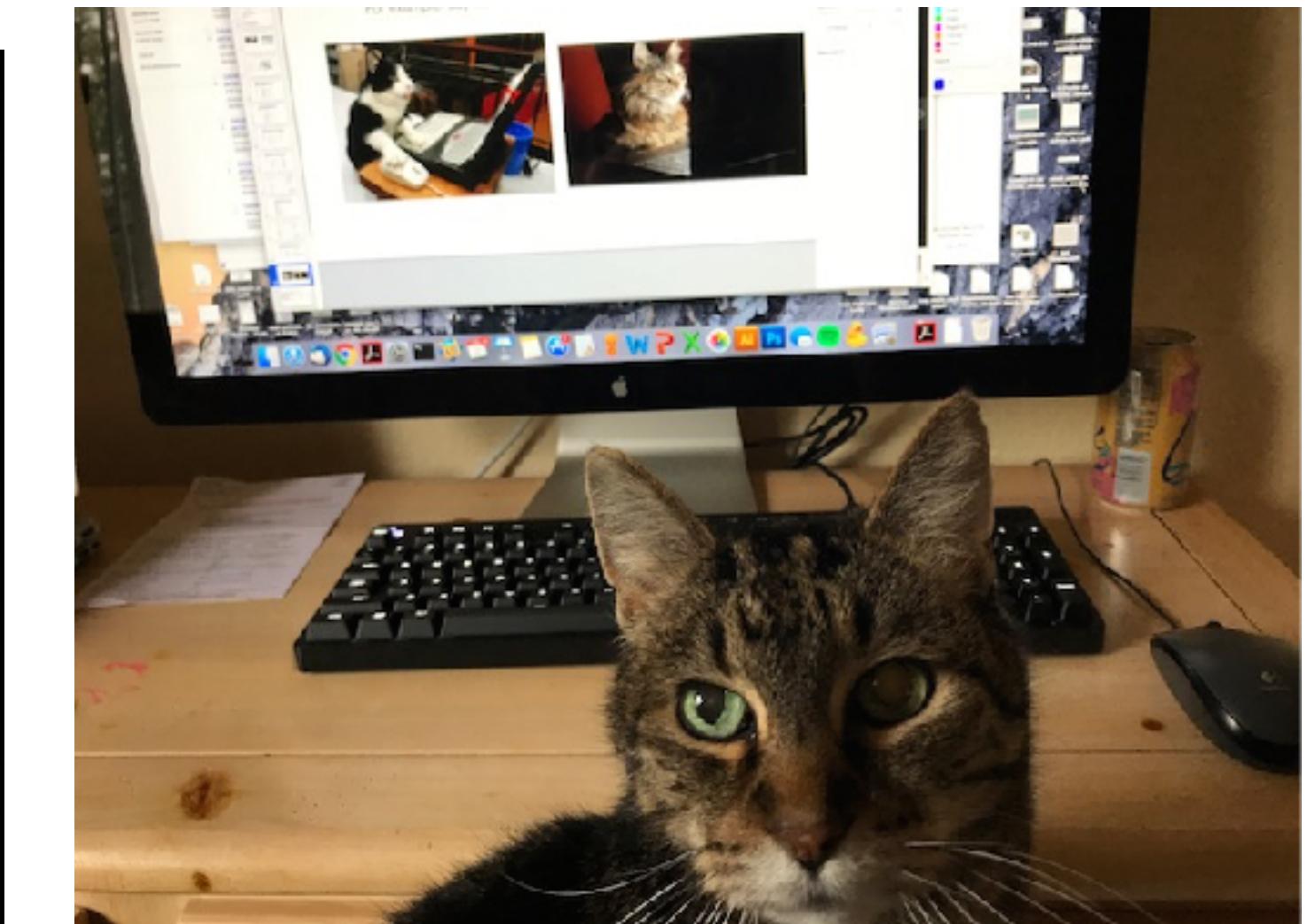
Items: 1 to 20 of 26

- [Camelus ferus MHC class II antigen \(Cafe-DQA\) gene, Cafe-DQA*02 allele](#)
1. 249 bp linear DNA
Accession: KT936421.1 GI: 1003311003
[GenBank](#) [FASTA](#) [Graphics](#)
- [Camelus ferus MHC class II antigen \(Cafe-DQA\) gene, Cafe-DQA*01 allele](#)
2. 249 bp linear DNA
Accession: KT936420.1 GI: 1003311001
[GenBank](#) [FASTA](#) [Graphics](#)
- [Camelus dromedarius MHC class II antigen \(Cadr-DQA\) gene, Cadr-DQA*01 allele](#)
3. partial cds
249 bp linear DNA
Accession: KT936419.1 GI: 1003310999
[GenBank](#) [FASTA](#) [Graphics](#)
- [Camelus bactrianus MHC class II antigen \(Caba-DQA\) gene, Caba-DQA*01 allele](#)
4. partial cds

← includes annotations
← just the sequences

There are often many paths to the same data

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



Kirby, 17 year old male cat

You could try to get the cat genome from the NCBI nucleotide db

<https://www.ncbi.nlm.nih.gov/nuccore/?term=Felis+catus>

NCBI Resources ▾ How To ▾ Sign in to NCBI

Nucleotide Nucleotide Felis catus Create alert Advanced Help

Species Summary ▾ 20 per page ▾ Sort by Default order ▾ Send: ▾ Filters: Manage Filters

Animals (355,888)
Plants (10)
Fungi (195)
Protists (38,091)
Bacteria (5,536)
Viruses (3,369)
Customize ...

Items: 1 to 20 of 403089 ←

Found 407775 nucleotide sequences. Nucleotide (403089) EST (1089) GSS (3597)

[Felis catus polycystic kidney disease 1-like \(PKD1\) gene, exons 21, 22, 23, 24, 29, 37, and 38](#)
1. 3,269 bp linear DNA
Accession: AH014595.2 GI: 1059791198
[GenBank](#) [FASTA](#) [Graphics](#)

[Felis catus isolate Cinnamon breed Abyssinian chromosome A1, Felis_catus_8.0, whole genome](#)
2. shotgun sequence
240,380,223 bp linear DNA
Accession: NC_018723.2 GI: 753572113
[GenBank](#) [FASTA](#) [Graphics](#)

[Felis catus isolate Cinnamon breed Abyssinian chromosome A2, Felis_catus_8.0, whole genome](#)
3. shotgun sequence
168,638,799 bp linear DNA
Accession: NC_018724.2 GI: 753572104
[GenBank](#) [FASTA](#) [Graphics](#)

[Felis catus isolate Cinnamon breed Abyssinian chromosome A3, Felis_catus_8.0, whole genome](#)
4. shotgun sequence
140,925,898 bp linear DNA
Accession: NC_018725.2 GI: 753572100
[GenBank](#) [FASTA](#) [Graphics](#)

Results by taxon

Top Organisms [Tree]
Felis catus (355407) ←
Hammonia hammondi (37729)
Bartonella henselae (1471)
Feline immunodeficiency virus (1391) ←
Yersinia pestis PY-47 (1070)
All other taxa (6021)
More...

Find related data

Database: Select

Find items

Search details

"Felis catus"[Organism] OR Felis catus[All Fields]

Search See more...

One good way to get the cat genome is via the Genome database

← → ⌂ Secure <https://www.ncbi.nlm.nih.gov/genome/?term=felis+catus>

NCBI Resources How To

Genome [Genome](#) [Create alert](#) [Limits](#) [Advanced](#)

Felis catus (domestic cat)
Reference genome: [Felis catus \(assembly Felis_catus_8.0\)](#)
Download sequences in FASTA format for [genome](#), [transcript](#), [protein](#)
Download genome annotation in [GFF](#), [GenBank](#) or [tabular](#) format
[BLAST](#) against *Felis catus* [genome](#)
All 2 genomes for species:
[Browse the list](#)
[Download sequence and annotation from RefSeq or GenBank](#)

Display Settings: [Overview](#) Send to: [▼](#)

[Organism Overview](#) ; [Genome Assembly and Annotation report \[2\]](#) ; [Organelle Annotation Report \[1\]](#) ID: 78

 ***Felis catus (domestic cat)***
domestic cat

Lineage: [Eukaryota\[2334\]](#); [Metazoa\[779\]](#); [Chordata\[332\]](#); [Craniata\[324\]](#); [Vertebrata\[324\]](#); [Euteleostomi\[319\]](#); [Mammalia\[136\]](#); [Eutheria\[131\]](#); [Laurasiatheria\[61\]](#); [Carnivora\[13\]](#); [Feliformia\[4\]](#); [Felidae\[4\]](#); [Felineae\[1\]](#); [Felis\[1\]](#); [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...](#)

There are actually 2 cat genome assemblies in NCBI

← → ⌂ Secure https://www.ncbi.nlm.nih.gov/genome/?term=felis+catus

NCBI Resources How To

Genome Genome felis catus[orgn]

Create alert Limits Advanced

Felis catus (domestic cat)
Reference genome: [Felis catus \(assembly Felis_catus_8.0\)](#)
Download sequences in FASTA format for [genome](#), [transcript](#), [protein](#)
Download genome annotation in [GFF](#), [GenBank](#) or [tabular](#) format
BLAST against Felis catus [genome](#)

All 2 genomes for species: [Browse the list](#) ←

Display Settings: Overview Send to: ID: 78

[Organism Overview](#) ; [Genome Assembly and Annotation report \[2\]](#) ; [Organelle Annotation Report \[1\]](#)

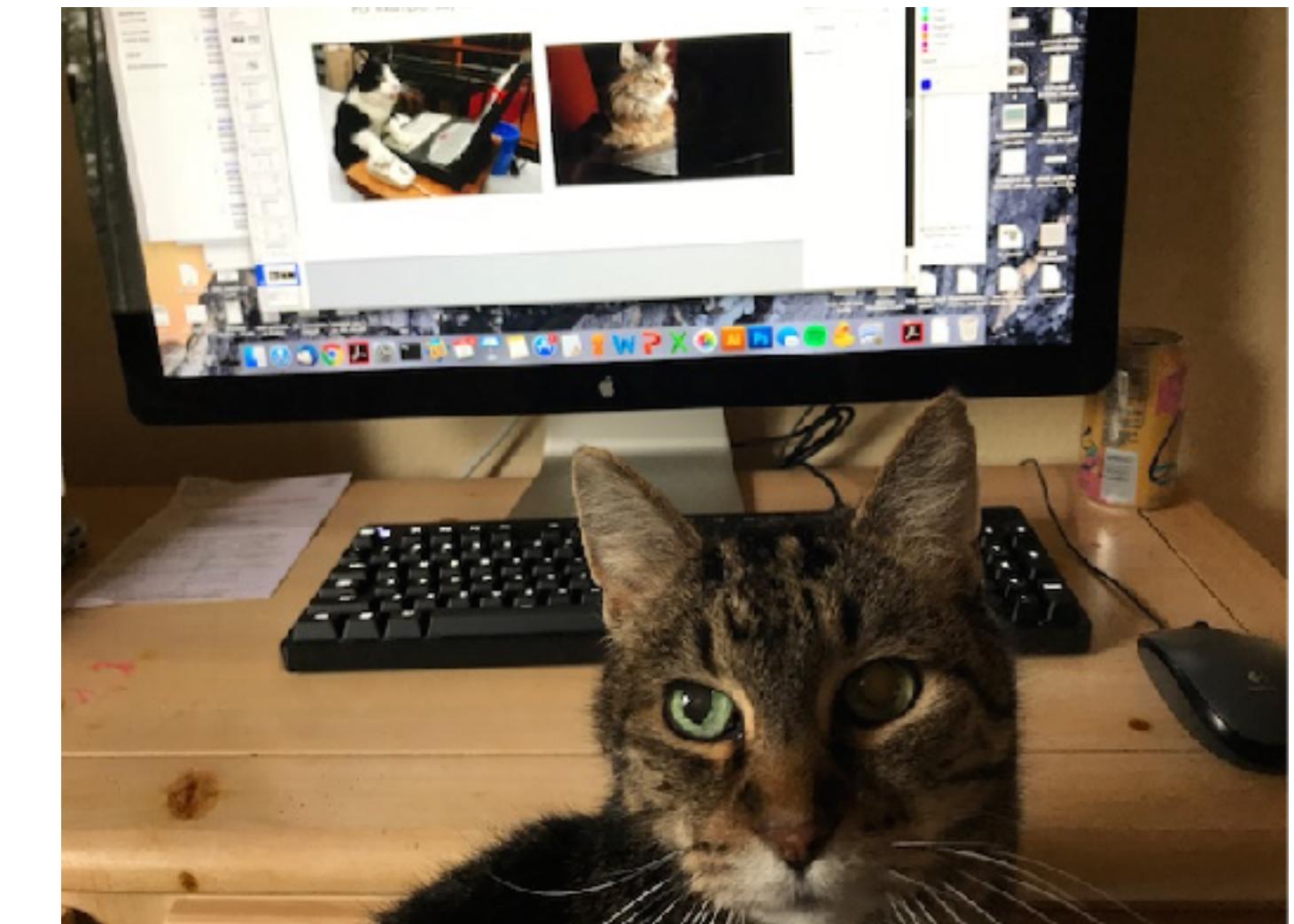
 **Felis catus (domestic cat)**
domestic cat

Lineage: [Eukaryota](#)[2334]; [Metazoa](#)[779]; [Chordata](#)[332]; [Craniata](#)[324]; [Vertebrata](#)[324]; [Euteleostomi](#)[319]; [Mammalia](#)[136]; [Eutheria](#)[131]; [Laurasiatheria](#)[61]; [Carnivora](#)[13]; [Feliformia](#)[4]; [Felidae](#)[4]; [Felineae](#)[1]; [Felis](#)[1]; [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...](#)

In reality, there are as many cat genomes as there are cats

Or maybe 2x as many...



Kirby, 17 year old male cat

There are 2 cat genome assemblies in NCBI

NCBI Resources How To

Assembly Assembly Advanced Browse by organism

Full Report 

catChrV17e

Organism name: [Felis catus \(domestic cat\)](#)
Infraspecific name: Breed: mixed
BioSample: [SAMN03000705](#)
Submitter: NHGRI/Genome Technology Branch
Date: 2009/03/16
Synonyms: felCat4
Assembly level: Chromosome
Genome representation: full
GenBank assembly accession: GCA_000003115.1 (latest)
RefSeq assembly accession: n/a
RefSeq assembly and GenBank assembly identical: n/a
WGS Project: [ACBE01](#)
IDs: 3328 [UID] 3328 [GenBank]

[History](#) (Show revision history)

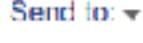
Comment
This assembly adds 0.9X coverage to the 1.9X coverage of the previous cat assembly (WGS project AANG00000000 of the Abyssinian cat "Cinnamon"). The 0.9X coverage comes from six other breeds and one wild cat from paired-end fosmid ABI 3730 ... [more](#)

Global statistics

Total sequence length	3,160,286,939
Total assembly gap length	1,169,667,819
Gaps between scaffolds	381
Number of scaffolds	104,434
Scaffold N50	16,304,361
Scaffold L50	47
Number of contigs	604,920
Contig N50	4,577
Contig L50	122,526
Total number of chromosomes and plasmids	19

NCBI Resources How To

Assembly Assembly Advanced Browse by organism

Full Report 

Felis_catus_8.0

Organism name: [Felis catus \(domestic cat\)](#)
Infraspecific name: Breed: Abyssinian
Isolate: Cinnamon
Sex: female
BioSample: SAMNC02953640
Submitter: International Cat Genome Sequencing Consortium
Date: 2014/11/07
Synonyms: felCat8
Assembly level: Chromosome
Genome representation: full
RefSeq category: representative genome
GenBank assembly accession: GCA_000181335.3 (latest)
RefSeq assembly accession: GCF_000181335.2 (latest)
RefSeq assembly and GenBank assembly identical: no (hide details)

- Only in RefSeq: chromosome MT.
- Data displayed for RefSeq version

WGS Project: [AANGC3](#)
Assembly method: CABOG v. 6.2; MaSuRCA assembler v. 8.0; GAA v. 1.0
Genome coverage: 2x Sanger; 14x 454; 20x Illumina
Sequencing technology: Sanger; 454 Titanium; Illumina
IDs: 250841 [UID] 1373248 [GenBank] 1513828 [RefSeq]

[History](#) (Show revision history)

Comment
A female Abyssinian cat named Cinnamon kept by Dr. Kristina Narfstrom at the University of Missouri was used as the DNA source for all sequencing reads. From this source the Broad Institute and Agencourt have generated 6.7M plasmid and ... [more](#)

Global statistics

Total sequence length	2,641,342,268
Total assembly gap length	41,625,436
Gaps between scaffolds	303
Number of scaffolds	267,928
Scaffold N50	18,072,971
Scaffold L50	45
Number of contigs	367,572
Contig N50	45,169
Contig L50	16,262
Total number of chromosomes and plasmids	20

There is often not 1 obviously 'best' version of what you're looking for

You could also get at the cat genome via the Taxonomy database

Secure | https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi

NCBI Taxonomy Browser

Search for: Felis catus as complete name lock Go

Display 3 levels using filter: none

Felis catus

Taxonomy ID: 9685
Genbank common name: domestic cat
Inherited blast name: carnivores
Rank: species
Genetic code: Translation table 1 (Standard)
Mitochondrial genetic code: Translation table 2 (Vertebrate Mitochondrial)
Other names:
synonym: **Felis silvestris catus**
synonym: **Felis domesticus**
common name: **cats**
common name: **cat**
includes: **Korat cats**
authority: **Felis catus Linnaeus, 1758**

[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	355,415
Nucleotide EST	919
Nucleotide GSS	3,107
Protein	55,602
Structure	10
Genome	1
Popset	172
SNP	3,611,275
GEO Datasets	103
PubMed Central	1,706
Gene	39,815
SRA Experiments	406
Probe	2,877
Assembly	4
Bio Project	39
Bio Sample	396
Bio Systems	495
Clone DB	239,767
PubChem BioAssay	1,097
Protein Clusters	12
Taxonomy	1

You can go up the taxonomic tree in the Taxonomy db

Secure | https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi

NCBI Taxonomy Browser

Search for: Felis catus as complete name lock Go

Display 3 levels using filter: none

Felis catus

Taxonomy ID: 9685
Genbank common name: domestic cat
Inherited blast name: carnivores
Rank: species
Genetic code: Translation table 1 (Standard)
Mitochondrial genetic code: Translation table 2 (Vertebrate Mitochondrial)
Other names:
synonym: **Felis silvestris catus**
synonym: **Felis domesticus**
common name: cats
common name: cat
includes: Korat cats
authority: **Felis catus Linnaeus, 1758**

[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	355,415
Nucleotide EST	919
Nucleotide GSS	3,107
Protein	55,602
Structure	10
Genome	1
Popset	172
SNP	3,611,275
GEO Datasets	103
PubMed Central	1,706
Gene	39,815
SRA Experiments	406
Probe	2,877
Assembly	4
Bio Project	39
Bio Sample	396
Bio Systems	495
Clone DB	239,767
PubChem BioAssay	1,097
Protein Clusters	12
Taxonomy	1

You can go up the taxonomic tree in the Taxonomy db

NCBI Taxonomy Browser

Search for: as complete name lock Go

Display 3 levels using filter: none

Nucleotide Nucleotide EST Nucleotide GSS Protein Structure Genome Popset SNP
 Domains GEO Datasets UniGene PubMed Central Gene HomoloGene SRA Experiments Probe
 Assembly LinkOut BLAST TRACE Host Viral Host Bio Project Bio Sample
 Bio Systems Clone DB dbVar GEO Profiles PubChem BioAssay Protein Clusters

[Lineage \(full\)](#): root; cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Amniota; Mammalia; Theria; Eutheria; Boreoeutheria; Laurasiatheria; Carnivora; Feliformia

- [Felidae](#) (cat family) 36 Click on organism name to get more information.
 - [Acinonychiae](#) 1
 - [Acinonyx](#) 1
 - [Acinonyx jubatus](#) (cheetah) 1 ←
 - [Felinae](#) 28
 - [Caracal](#) 1
 - [Caracal caracal](#) 1
 - [Catopuma](#) 2
 - [Catopuma badia](#) (bay cat) 1
 - [Catopuma temminckii](#) (Asiatic golden cat) 1
 - [Felis](#) 5
 - [Felis catus](#) (domestic cat) 1
 - [Felis chaus](#) (jungle cat) 1
 - [Felis margarita](#) (sand cat) 1
 - [Felis nigripes](#) (black-footed cat) 1
 - [Felis silvestris](#) (wild cat) 1
 - [Felis sp.](#)
 - [Felis sp. NG192](#)
 - [Leopardus](#) 7
 - [Leopardus colocolo](#) (Colocolo) 1
 - [Leopardus geoffroyi](#) (Geoffroy's cat) 1
 - [Leopardus guigna](#) (Kodkod) 1
 - [Leopardus guttulus](#) (Southern oncilla) 1
 - [Leopardus jacobita](#) (Andean mountain cat) 1
 - [Leopardus pardalis](#) (ocelot) 1
 - [Leopardus tigrinus](#) (little spotted cat) 1
 - [Leopardus wiedii](#) (margay) 1
 - [Leptailurus](#) 1
 - [Leptailurus serval](#) (serval) 1
 - [Leptailurus serval x Caracal caracal](#)
 - [Lynx](#) 4
 - [Lynx canadensis](#) (Canada lynx) 1
 - [Lynx lynx](#) (Eurasian lynx) 1
 - [Lynx pardinus](#) (Spanish lynx) 1
 - [Lynx rufus](#) (bobcat) 1
 - [Otocolobus](#) 1

You can go up the taxonomic tree in the Taxonomy db

The screenshot shows the NCBI Genome page for the cheetah (*Acinonyx jubatus*). The search bar at the top contains the query "txid32536[Organism:exp]". The main content area displays the following information:

Acinonyx jubatus (cheetah)
Representative genome: [Acinonyx jubatus \(assembly aciJub1\)](#)
Download sequences in FASTA format for [genome](#), [transcript](#), [protein](#)
Download genome annotation in [GFF](#), [GenBank](#) or [tabular](#) format
[BLAST against Acinonyx jubatus genome](#)

Display Settings: ▾ Overview Send to: ▾ ID: 14623

[Organism Overview](#) ; [Organelle Annotation Report \[1\]](#)

Acinonyx jubatus (cheetah)

[Acinonyx jubatus overview](#)

Lineage: [Eukaryota](#)[2334]; [Metazoa](#)[779]; [Chordata](#)[332]; [Craniata](#)[324]; [Vertebrata](#)[324]; [Euteleostomi](#)[319]; [Mammalia](#)[136]; [Eutheria](#)[131]; [Laurasiatheria](#)[61]; [Carnivora](#)[13]; [Feliformia](#)[4]; [Felidae](#)[4]; [Acinonychinae](#)[1]; [Acinonyx](#)[1]; [Acinonyx jubatus](#)[1]

Summary

Submitter:	Saint Petersburg State University
Assembly level:	Scaffold
Environment:	OptimumTemperature: C
Assembly:	GCA_001443585.1 aciJub1 scaffolds: 14,383 contigs: 170,171 N50: 35,120 L50: 19,059
BioProjects:	PRJNA305718, PRJNA297632
Whole Genome Shotgun (WGS):	INSDC: LLWD00000000.1
Statistics:	total length (Mb): 2372.55 protein count: 27284 GC%: 41.4
NCBI Annotation Release:	100

Publications

You need not rely on your browser to download data

Secure | https://www.ncbi.nlm.nih.gov/genome/?term=felis+catus

NCBI Resources How To

Genome

Genome | felis catus[orgn]

Create alert Limits Advanced

Felis catus (domestic cat)

Reference genome: [Felis catus \(assembly Felis_catus_8.0\)](#)

Download sequences in FASTA format for [genome](#), [transcript](#), [protein](#)

Download genome annotation in [GFF](#), [GenBank](#) or tabular format

BLAST against [Felis catus genome](#)

All 2 genomes for species:

Browse the list

Download sequence and annotation from [RefSeq](#) or [GenBank](#)

Display Settings: Overview

Send to: ▾

Organism Overview ; Genome Assembly and Annotation report [2] ; Organelle Annotation Report [1]

ID: 78

Felis catus (domestic cat)

domestic cat

Lineage: Eukaryota[2334]; Metazoa[779]; Chordata[332]; Craniata[324]; Vertebrata[324]; Euteleostomi[319]; Mammalia[136]; Eutheria[131]; Laurasiatheria[61]; Carnivora[13]; Feliformia[4]; Felidae[4]; Felinae[1]; Felis[1]; 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...](#)

FTP links

ftp://ftp.ncbi.nlm.nih.gov/genomes/

Index of /genomes/

Name	Size	Date Modified
[parent directory]		6/2/17, 9:57:00 AM
ASSEMBLY_REPORTS/		1/8/15, 5:00:00 PM
Acanthisitta_chloris/		12/15/15, 5:00:00 PM
Acinonyx_jubatus/		6/16/15, 6:00:00 PM
Acromyrmex_echinatior/		3/14/16, 6:00:00 PM
Acropora_digitifera/		5/17/16, 6:00:00 PM
Acyrtosiphon_pisum/		2/6/17, 9:24:00 AM
Aedes_albopictus/		3/3/17, 8:44:00 AM
Aegilops_tauschii/		1/19/17, 2:14:00 PM
Aethina_tumida/		9/22/16, 6:00:00 PM
Agrilus_planipennis/		1/3/17, 9:24:00 AM
Ailuropoda_melanoleuca/		
...		

You can download data from the command line

This is often useful when you're working on a server.



Secure https://www.ncbi.nlm.nih.gov/genome/?term=felis+catus

NCBI Resources How To

Genome

Genome felis catus[orgn]

Create alert Limits Advanced

Felis catus (domestic cat)

Reference genome: **Felis catus (assembly Felis_catus_8.0)**

Download sequences in FASTA format for **genome, transcript, protein**
Download genome annotation in **GFF, GenBank or tabular format**
BLAST against Felis catus genome

All 2 genomes for species:

Browse the list
Download sequence and annotation from **RefSeq or GenBank**

Display Settings: Overview

Send to:

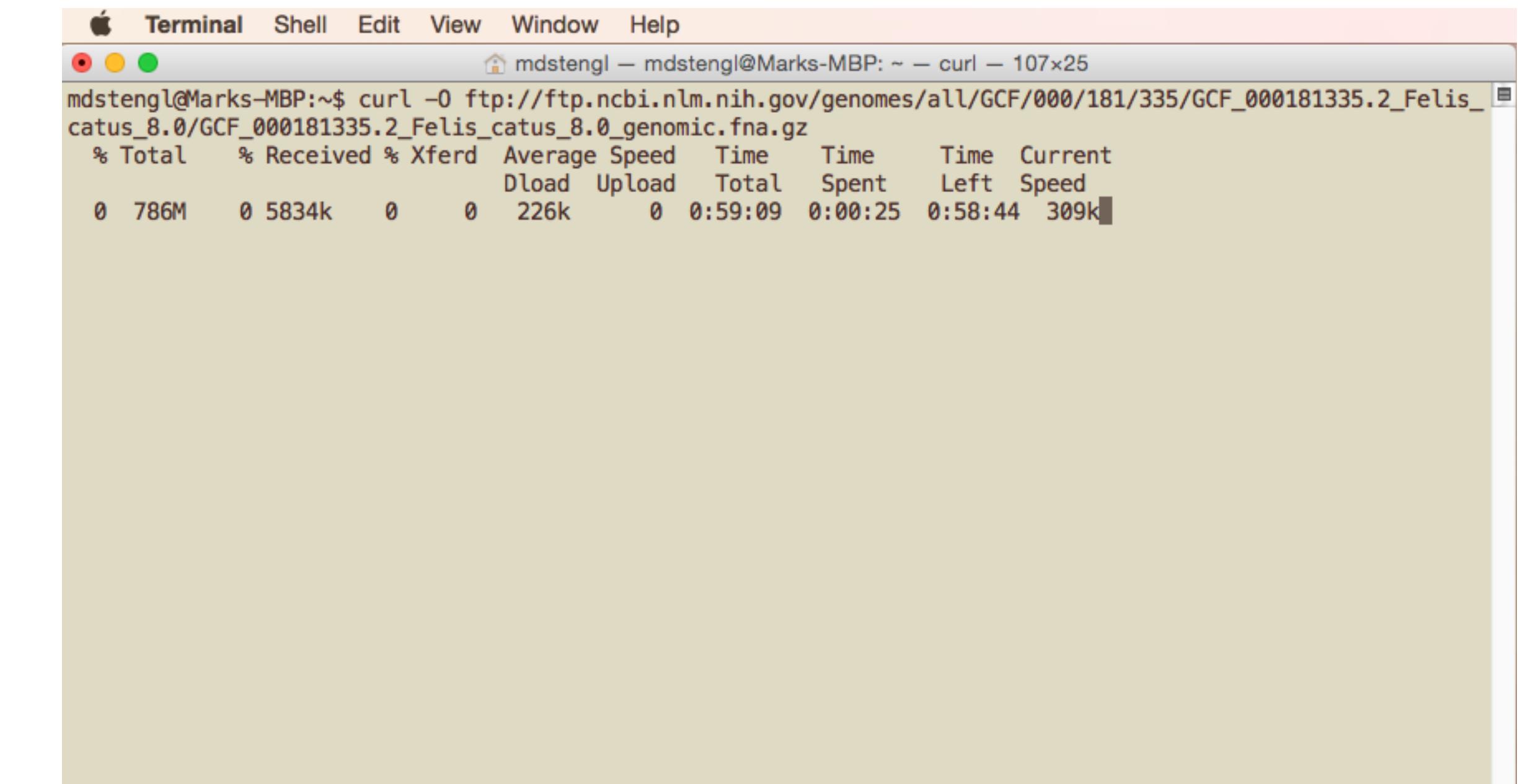
Organism Overview ; Genome Assembly and Annotation report [2] ; Organelle Annotation Report [1]

ID: 78

Felis catus (domestic cat)
domestic cat

Lineage: Eukaryota[2334]; Metazoa[779]; Chordata[332]; Craniata[324]; Vertebrata[324]; Euteleostomi[319]; Mammalia[136]; Eutheria[131]; Laurasiatheria[61]; Carnivora[13]; Feliformia[4]; Felidae[4]; Felinae[1]; Felis[1]; 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...](#)



```
Terminal Shell Edit View Window Help
mdstengl@Marks-MBP:~$ curl -O ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/181/335/GCF_000181335.2_Felis_catus_8.0/GCF_000181335.2_Felis_catus_8.0_genomic.fna.gz
% Total % Received % Xferd Average Speed Time Time Current
          Dload Upload Total Spent Left Speed
0 786M  0 5834k   0     0  226k      0  0:59:09  0:00:25  0:58:44  309k
```

curl is a file transfer utility built into Linux, MacOS

similar utilities exist for Windows

GUI-based software for file transfer

Cyberduck



Screenshot of the Cyberduck application interface showing an FTP connection to `ftp.ncbi.nlm.nih.gov`. The interface includes a toolbar with 'Connect to server', 'Action', 'Refresh', and 'Edit' buttons. The main window displays a file list with columns for 'Filename', 'Size', and 'Modified'. The list shows 46 files, including 'genomes', 'bioproject', 'biosample', 'snp', 'repository', 'pubchem', 'pathogen', 'hmm', 'geo', 'pubmed', 'refseq', 'sra', 'ncbi-asn1', 'genbank', 'eqtl', 'dbgap', 'pub', 'sequin', 'variation', and 'nist-immsa'. The 'genomes' file is highlighted with a blue selection bar. The bottom status bar indicates '46 Files'.

`ftp://ftp.ncbi.nlm.nih.gov/`



You can search and download from NCBI databases directly in Geneious

Geneious 10.0.7

Back Forward BLAST Workflows Align/Assemble Tree Primers Cloning Back Up Support Help

Sources

Local (84317, 18982 unread) Shared Databases Operations NCBI Gene Genome (20) Nucleotide NCBI Genome database (32,686 records) Protein PubMed SNP Structure Taxonomy UniProt

Download completed

Felis catus

Sequence Length	Name	Description
94,101,111	NC_018734	Felis catus isolate Cinnamon breed Abyssinian chromosome D3, Felis_catus_8.0, whole genome shotgun sequence
88,096,124	NC_018733	Felis catus isolate Cinnamon breed Abyssinian chromosome D2, Felis_catus_8.0, whole genome shotgun sequence
83,953,389	NC_018740	Felis catus isolate Cinnamon breed Abyssinian chromosome F2, Felis_catus_8.0, whole genome shotgun sequence
70,119,229	NC_018739	Felis catus isolate Cinnamon breed Abyssinian chromosome F1, Felis_catus_8.0, whole genome shotgun sequence
61,960,243	NC_018737	Felis catus isolate Cinnamon breed Abyssinian chromosome E2, Felis_catus_8.0, whole genome shotgun sequence
61,081,816	NC_018736	Felis catus isolate Cinnamon breed Abyssinian chromosome E1, Felis_catus_8.0, whole genome shotgun sequence
41,224,383	NC_018738	Felis catus isolate Cinnamon breed Abyssinian chromosome E3, Felis_catus_8.0, whole genome shotgun sequence
17,009	NC_001700	Felis catus mitochondrial, complete genome

Sequence View Annotations Dotplot (Self) Text View Fasta Nucleotide View Summary XML Linear

← → Extract R.C. Translate Add/Edit Annotation Allow Editing Annotate & Predict Save

contro region: CR

NC_001700
17,009 bp

The circular diagram illustrates the mitochondrial genome of Felis catus (NC_001700). The genome is 17,009 bp long and contains several genes: CYTB gene (green), CYTB CDS (yellow), ND6 gene (green), ND6 CDS (yellow), ND5 gene (green), ND5 CDS (yellow), ND4 gene (green), ND4 CDS (yellow), ND4L gene (green), ND4L CDS (yellow), ND3 gene (green), ND3 CDS (yellow), COX3 gene (green), COX3 CDS (yellow), ATP6 gene (green), ATP6 CDS (yellow), ATP8 gene (green), ATP8 CDS (yellow), COX2 gene (green), COX2 CDS (yellow), COX1 gene (green), COX1 CDS (yellow), ND1 gene (green), ND1 CDS (yellow), ND2 gene (green), ND2 CDS (yellow), I-rRNA (red), and 5S rRNA (yellow). The diagram also shows the control region (CR) and the replication origin (rep origin). Various regions are labeled with their lengths: 17,009, 16,000, 15,000, 14,000, 13,000, 12,000, 11,000, 10,000, 9,000, 8,000, 7,000, 6,000, 5,000, 4,000, 3,000, 2,000, 1,000, and 300.

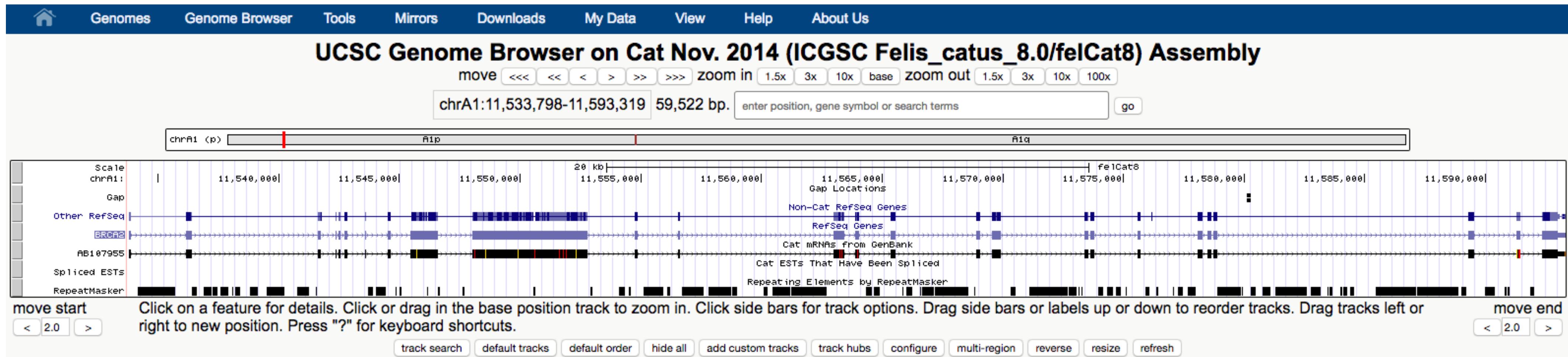
Updating search index: 2 items

+ -

Using 245 / 7996 MB memory

Alt click on a sequence position or annotation, or select a region to zoom in. Alt-shift click to zoom out.

Genome browsers, like Ensembl and UCSC, offer additional functionality



Finally, there's absolutely nothing wrong with using Google

cat genome

All Shopping Images News Videos More Settings Tools

About 11,500,000 results (0.60 seconds)

Specifically, the signatures of selection in the domestic cat genome are linked to genes associated with gene knockout models affecting memory, fear-conditioning behavior, and stimulus-reward learning, and potentially point to the processes by which cats became domesticated. Dec 2, 2014

Comparative analysis of the domestic cat genome reveals genetic ...
www.pnas.org/content/111/48/17230.abstract

The cat genome - NCBI
https://www.ncbi.nlm.nih.gov/genome?term=felis%20catus
1. Felis catus domestic cat Kingdom: Eukaryota Chromosomes: 19 Organelles: 1 Genome ID: 78.

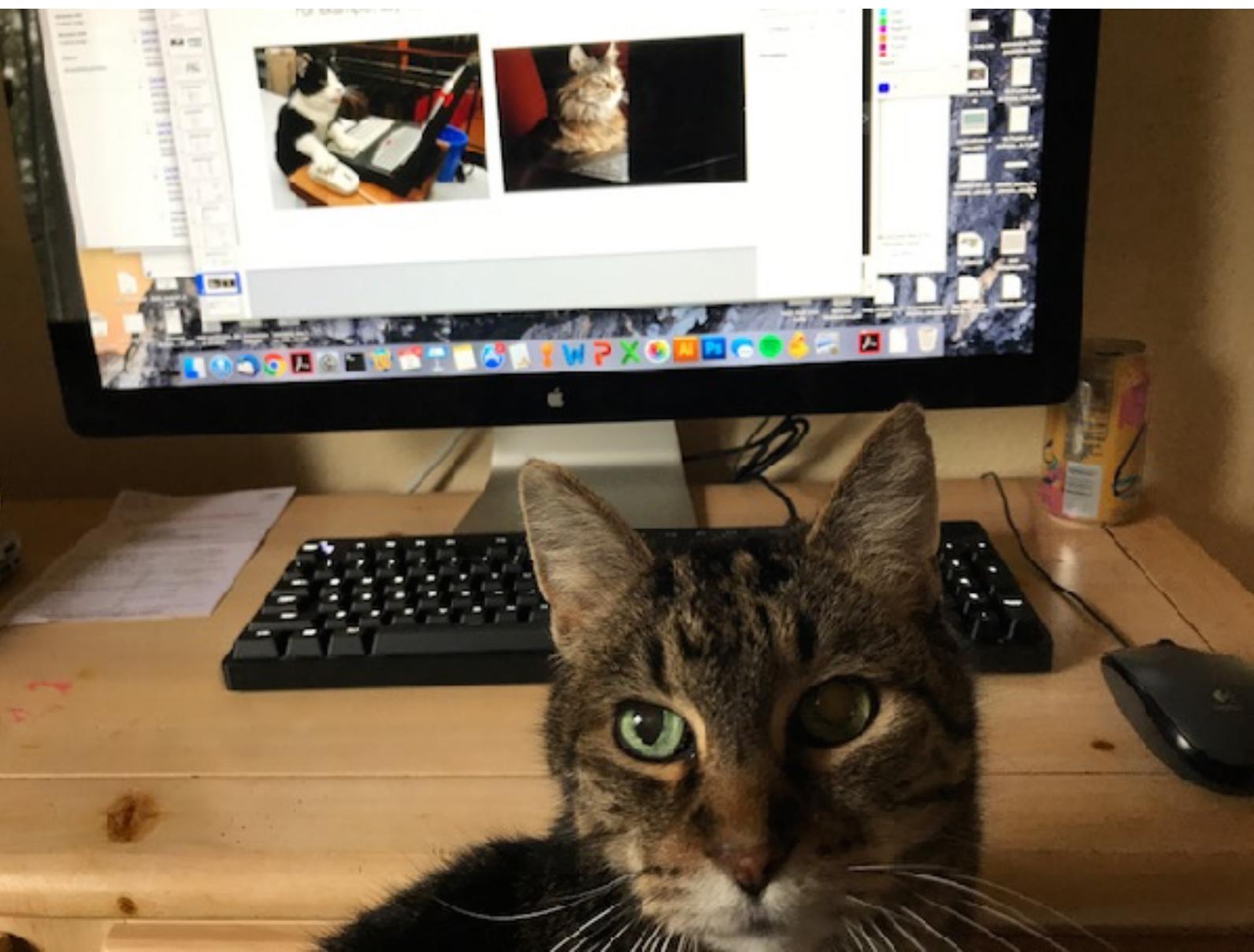
Comparative analysis of the domestic cat genome reveals genetic ...
www.pnas.org/content/111/48/17230.abstract ▾
by MJ Montague - 2014 - Cited by 67 - Related articles
Dec 2, 2014 - Specifically, the signatures of selection in the domestic cat genome are linked to genes associated with gene knockout models affecting memory, fear-conditioning behavior, and stimulus-reward learning, and potentially point to the processes by which cats became domesticated.

Felis catus - Ensembl genome browser 89
www.ensembl.org/Felis_catus/info/Index ▾
What's New in Cat release 89. Microarray Probe Mapping ... Genome assembly: Felis_catus_6.2 (GCA_000181335.2) ... assemblies. CAT (Ensembl release 67).

'I can haz genomes': cats claw their way into genetics : Nature News ...
www.nature.com/news/i-can-haz-genomes-cats-claw-their-way-into-genetics-1.16708 ▾
Jan 14, 2015 - The cat genome is out of the bag, and has already helped to pinpoint a gene involved in kidney disease. ... The work will benefit both humans and felines, the researchers say, by mapping the mutations underlying conditions that afflict the two species, such as kidney disease.

The Cat Genome Sequence - Lyons Feline & Comparative Genetics
felinegenetics.missouri.edu/feline-genome-project-2/the-cat-genome-sequence ▾
The Cat Genome Sequence. NCBI - Cat Genome Resources http://www.ncbi.nlm.nih.gov/projects/genome/guide/cat/ · Science Daily - Domestic Cat Genome ...

Questions?



Kirby in 2000, wondering where his GenBank CDROMs are