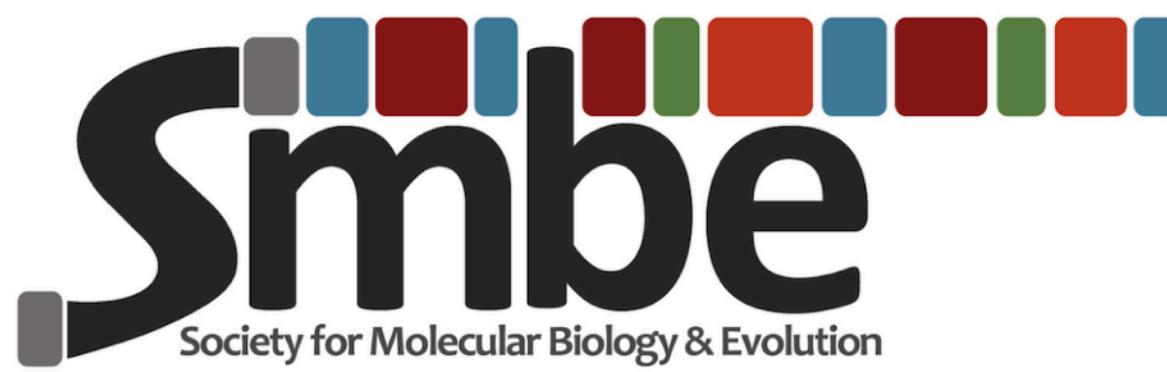


An overview of bioinformatics databases and online resources: what they are and how to access them

Mark Stenglein



SMBE Regional Workshop
on Computational Biology
in Todos Santos, Mexico



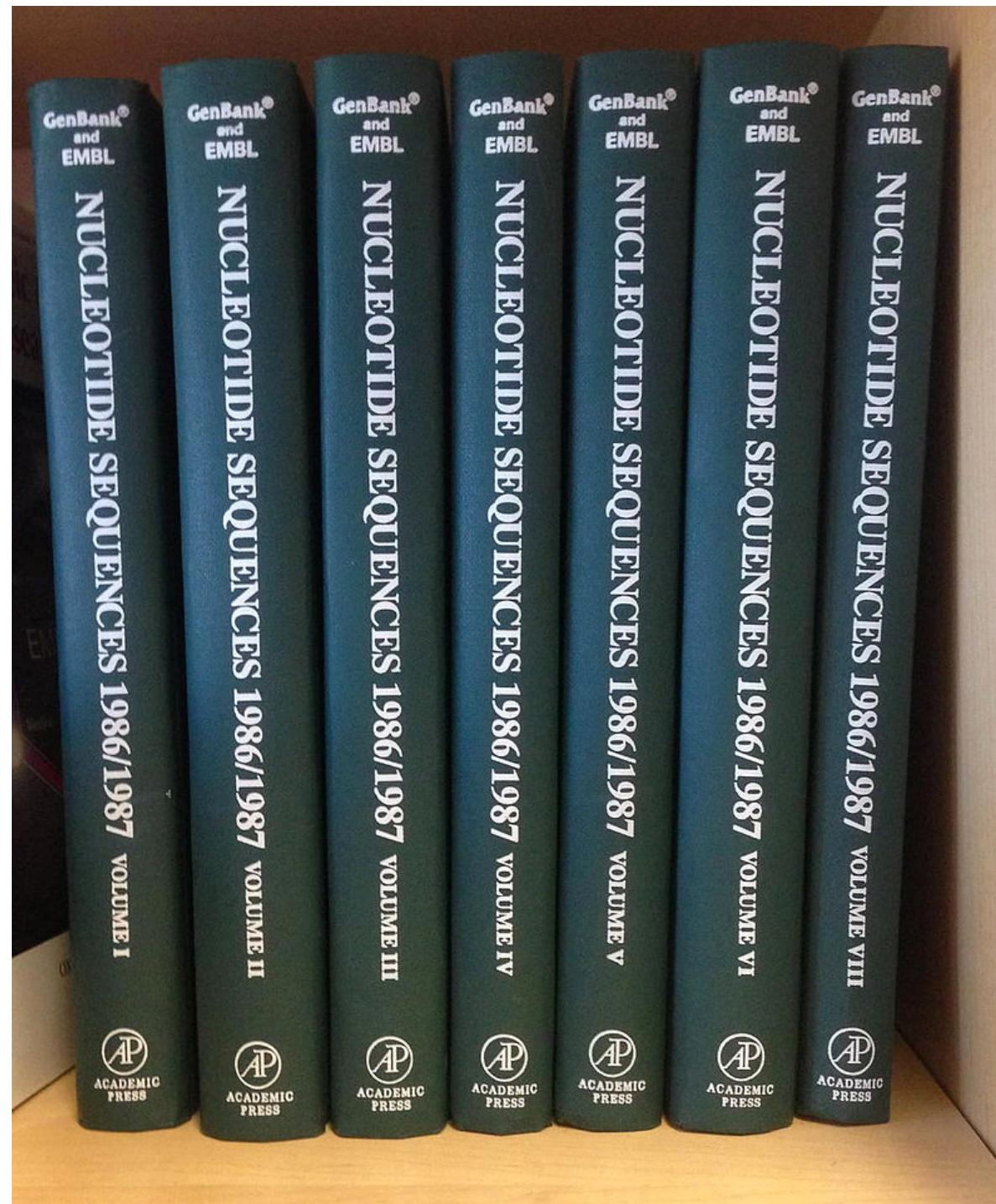
There are an overwhelming number of databases and other online resources, which often have overlapping content and purpose

The annual Database Nucleic Acids Research issue can be a good resource



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

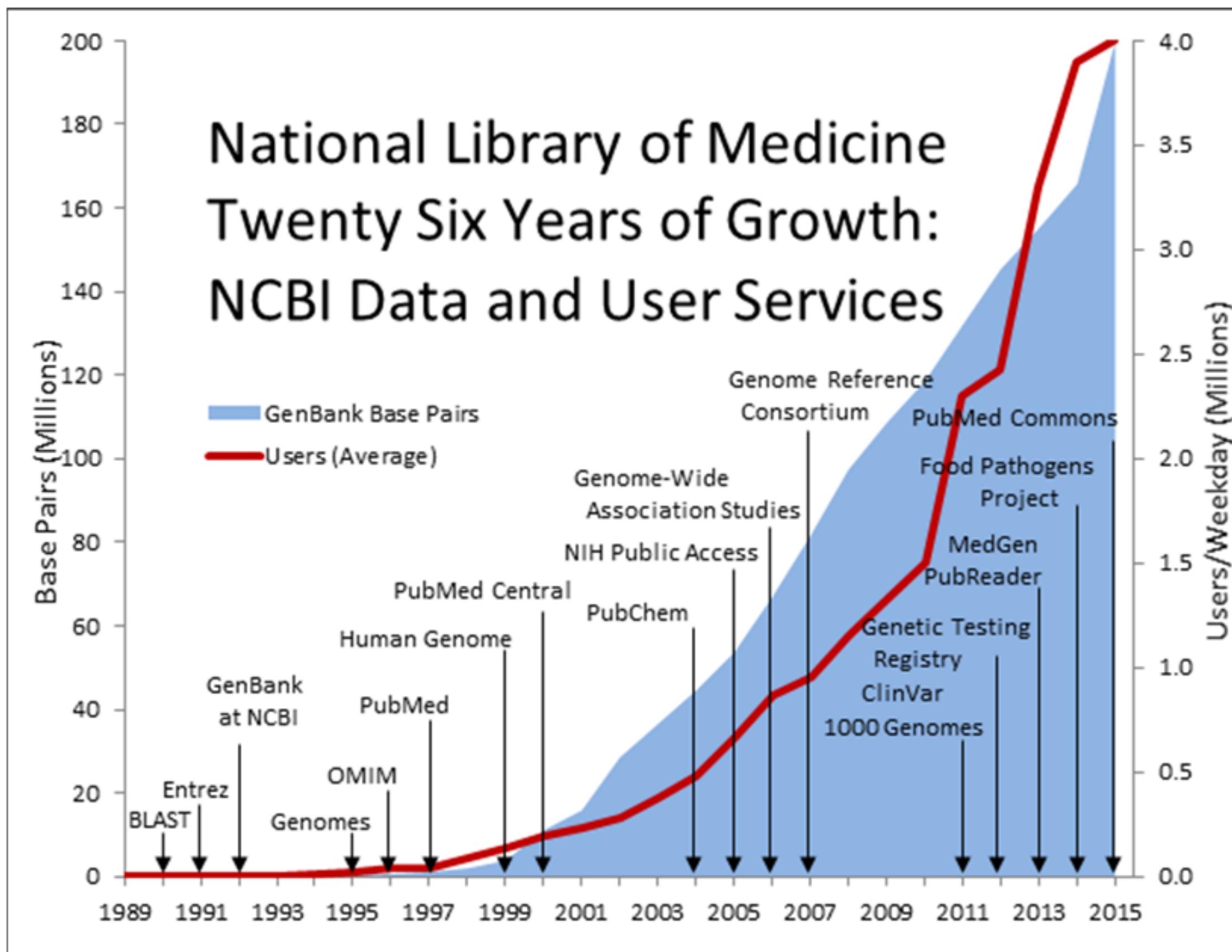


>200,000,000 sequences

BOVCHYMOA				NUCLEOTIDE SEQUENCES 1984						
SITES:	key	site	span	description	key	site	span	description		
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									
SEQUENCE:	1275	bp	293	a	391	c	336	g	255	t
1	ccgcctggacc	cagatccaag	atggagggtgc	tgttgtgtgt	acttgcgtgtc	ttcgctctct	cccaggccgc	ttagatcaccc	aggatccccctc	tgtacaaaaagg
101	caagtctctg	aggaaaggccgc	tgaaggaggta	tggcccttcgt	gaggacttcc	tgcagaaaaca	gcagttatggc	atcaggcgcga	agtactcccg	tttcggggggag
201	gtggccagcg	tgcctttctgac	caactaacctg	gatagtcaagt	actttgggaa	gatctacccctc	gggacccccgc	cccaggagtt	caccgtgtgt	tttgacactgt
301	gtccctctga	cttctgggta	ccctcttatct	actgcgaagag	caatgcctgc	aaaaaccacc	agcgttctcg	cccggaaaaag	tctgtccaccc	tccagaacact
401	gggcaagccc	ctgtctatacc	actacggggac	aggcagcatg	cagggccatcc	tgggttatgt	caccgtcaact	gttctccaaaca	tttcgcacat	ccagcggaaa
501	gtaggccctga	gcaccccaagg	gccccggggac	gttccatct	atgcggaaatt	cgacgggatc	ctggggatgg	cttacccccc	gttcgcctca	gagttactcg
601	taccctgtt	tgacaaacatg	atgcacaggc	acctgtgtgc	ccaaagacccgt	tttctcggtt	acatggacag	gaatggccag	gagagcatgc	tcaacgttgc
701	ggccatcgac	ccgtgtcaact	acacagggtc	cttcgtactgg	gtggccgtgt	cgtgtcgac	gtacgtggac	tttactgtgg	acatgttcac	catcgggggt
801	gtggttgtgg	cttgcgtgggg	tggctgtca	gcacatctgg	acacgggcac	cttcaaaatgt	gtcggggccca	gcacgcacat	cttcaacatc	cagcggggcc
901	ttggggccac	acagaacccag	tacgtatgtt	tttacatcg	ctgcgcacac	ctgaggctaca	tgcccaactgt	ggtttttgag	atcaatggca	aaatgtaccc
1001	actgacccccc	tccgcctata	ccagccagg	ccagggtttc	tgttaccagg	gtttccagag	tgaaaatcat	tcccaagaaaat	ggatccctgg	ggatgttttt
1101	atccgagagt	attacacgt	ctttgacagg	gccaacaacc	tcgtgggggt	ggccaaagcc	atctgtatca	atcgtgacc	aagaaacctca	ctgtccccac
1201	acacccgtcac	acacacatgc	acacatgtac	atggcacatg	tgcacacacaca	cagatpaggt	ttcccaagaccc	aagct		

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

The NCBI was created in 1987 by the US government



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

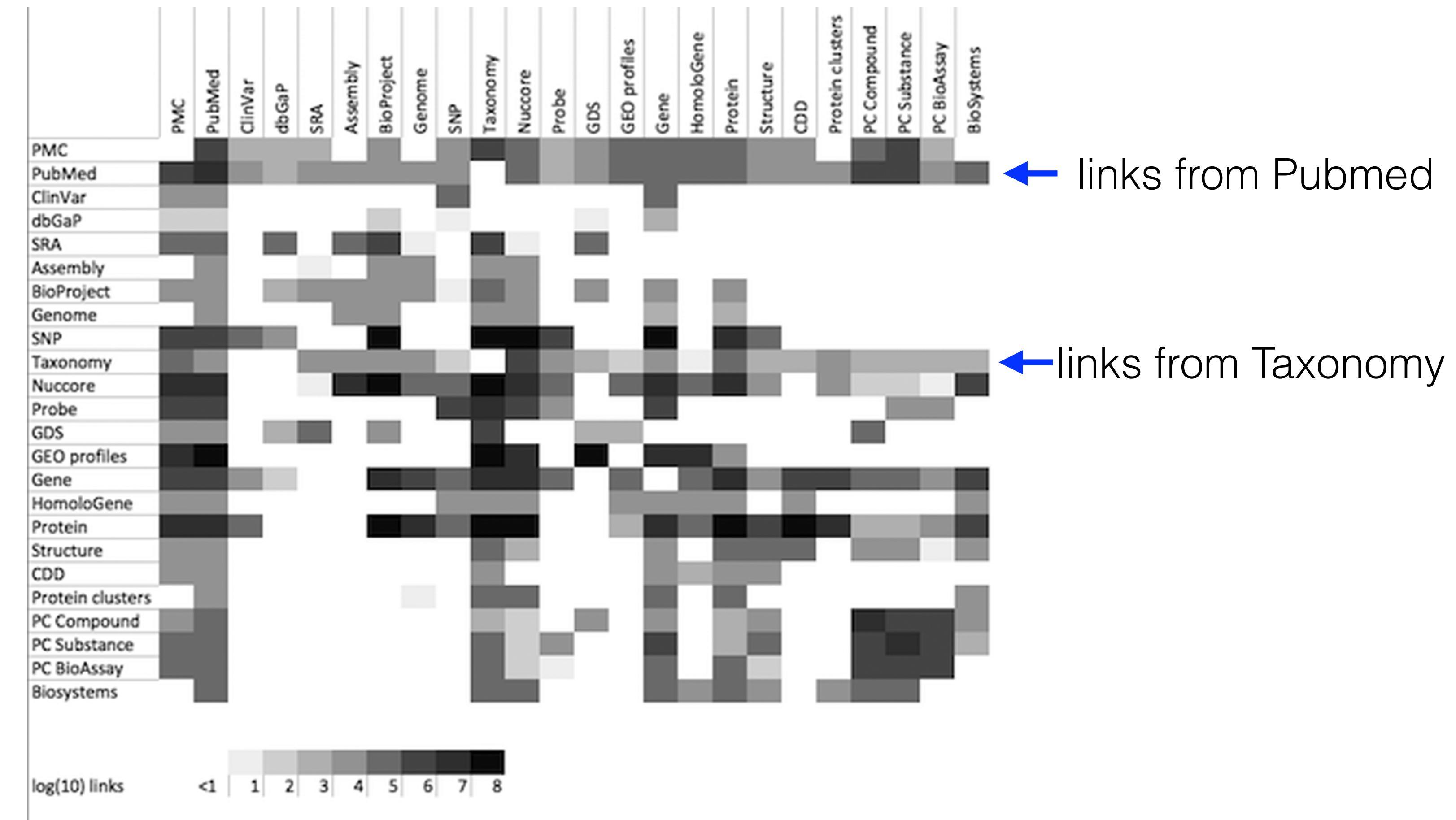
image: NIH/NLM

<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



Get nucleotide sequences associated with Dan's papers

NCBI Resources How To

PubMed Sloan DB

Create RSS Create alert Advanced Help

markstenglein My NCBI Sign Out

Article types Format: Summary Sort by: Publication Date Per page: 20 Send to Filters: Manage Filters

Search results Items: 1 to 20 of 50 << First < Prev Page 1 of 3 Next > Last >>

Text availability Abstract Free full text Full text

Publication dates 5 years 10 years Custom range...

Species Humans Other Animals

[Clear all](#) [Show additional filters](#)

[Detecting Rare Mutations and DNA Damage with Sequencing-Based Methods.](#)
1. **Sloan DB**, Broz AK, Sharbrough J, Wu Z.
Trends Biotechnol. 2018 Mar 14. pii: S0167-7799(18)30076-3. doi: 10.1016/j.tibtech.2018.02.009. [Epub ahead of print]
Review.
PMID: 29550161 [Similar articles](#)

[Correction of persistent errors in Arabidopsis reference mitochondrial genomes.](#)
2. **Sloan DB**, Wu Z, Sharbrough J.
Plant Cell. 2018 Mar 8. pii: tpc.00024.2018. doi: 10.1105/tpc.18.00024. [Epub ahead of print] No abstract available.
PMID: 29519893 [Free Article](#) [Similar articles](#)

[Mitochondrial Retroprocessing Promoted Functional Transfers of rpl5 to the Nucleus in Grasses.](#)
3. Wu Z, **Sloan DB**, Brown CW, Rosenblueth M, Palmer JD, Ong HC.
Mol Biol Evol. 2017 Sep 1;34(9):2340-2354. doi: 10.1093/molbev/msx170.
PMID: 28541477 [Similar articles](#)

[Nuclear and mitochondrial RNA editing systems have opposite effects on protein diversity.](#)
4. **Sloan DB**.
Biol Lett. 2017 Aug;13(8). pii: 20170314. doi: 10.1098/rsbl.2017.0314.
PMID: 28855414 [Similar articles](#)

[The Mitonuclear Dimension of Neanderthal and Denisovan Ancestry in Modern Human Genomes.](#)
5. Sharbrough J, Havird JC, Noe GR, Warren JM, **Sloan DB**.
Genome Biol Evol. 2017 Jun 1;9(6):1567-1581. doi: 10.1093/gbe/evx114.
PMID: 28854627 [Free PMC Article](#) [Similar articles](#)

[The on-again, off-again relationship between mitochondrial genomes and species boundaries.](#)
6. **Sloan DB**, Havird JC, Sharbrough J.
Mol Ecol. 2017 Apr;26(8):2212-2236. doi: 10.1111/mec.13959. Epub 2017 Jan 27. Review.
PMID: 27997046 [Similar articles](#)

Sort by: Best match Most recent

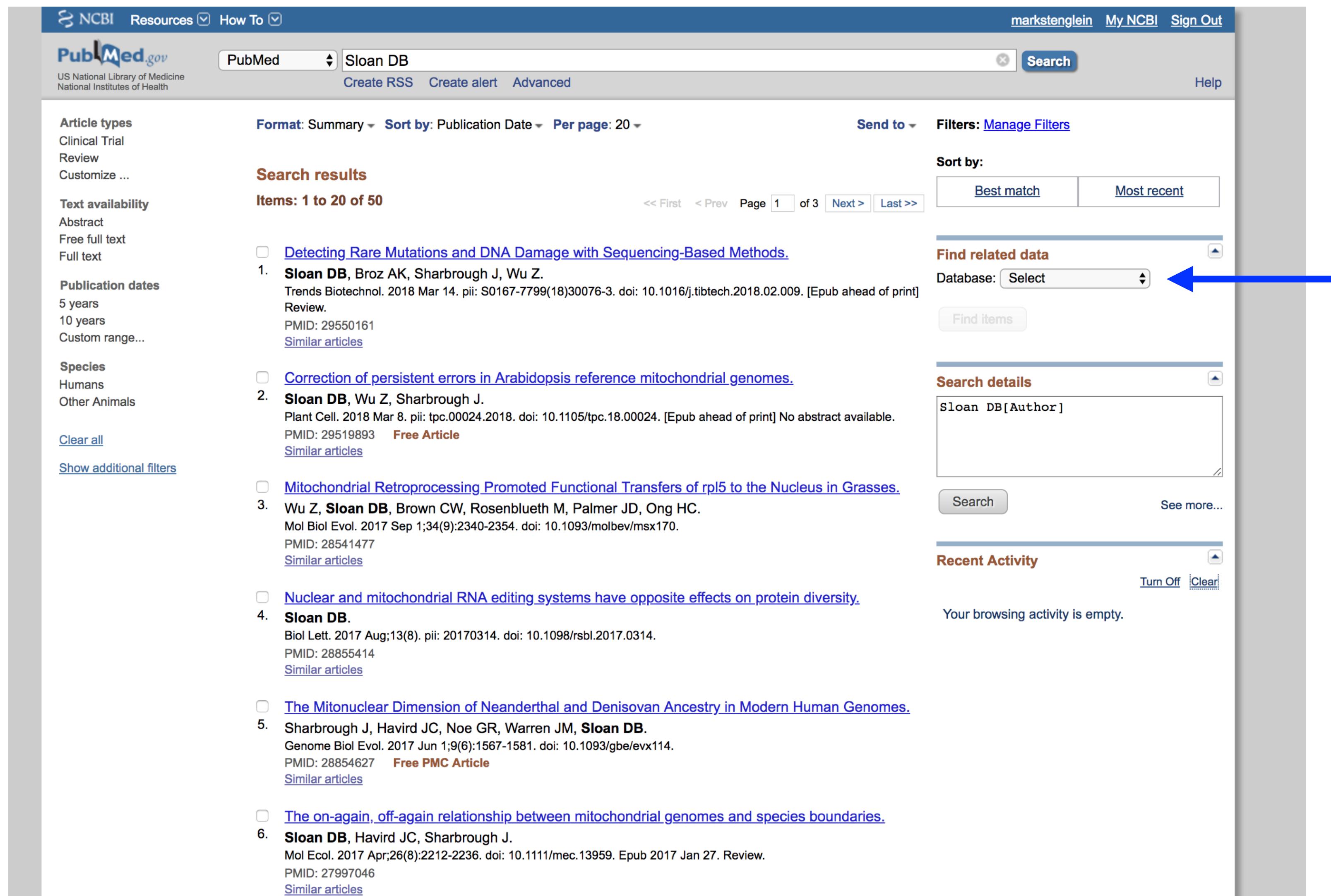
Find related data Database: Select

Find items

Search details Sloan DB[Author]

Search See more...

Recent Activity Turn Off Clear Your browsing activity is empty.



Get nucleotide sequences associated with Dan's publications

NCBI Resources How To

PubMed Sloan DB

Create RSS Create alert Advanced Help

markstenglein My NCBI Sign Out

Article types Format: Summary Sort by: Publication Date Per page: 20 Send to Filters: Manage Filters

Clinical Trial Review Customize ...

Text availability Abstract Free full text Full text

Publication dates 5 years 10 years Custom range...

Species Humans Other Animals

[Clear all](#) [Show additional filters](#)

Search results Items: 1 to 20 of 50 << First < Prev Page 1 of 3 Next > Last >>

1. [Detecting Rare Mutations and DNA Damage with Sequencing-Based Methods.](#) Sloan DB, Broz AK, Sharbrough J, Wu Z. Trends Biotechnol. 2018 Mar 14. pii: S0167-7799(18)30076-3. doi: 10.1016/j.tibtech.2018.02.009. [Epub ahead of print] Review. PMID: 29550161 [Similar articles](#)

2. [Correction of persistent errors in Arabidopsis reference mitochondrial genomes.](#) Sloan DB, Wu Z, Sharbrough J. Plant Cell. 2018 Mar 8. pii: tpc.00024.2018. doi: 10.1105/tpc.18.00024. [Epub ahead of print] No abstract available. PMID: 29519893 [Free Article](#) [Similar articles](#)

3. [Mitochondrial Retroprocessing Promoted Functional Transfers of rpl5 to the Nucleus in Grasses.](#) Wu Z, Sloan DB, Brown CW, Rosenblueth M, Palmer JD, Ong HC. Mol Biol Evol. 2017 Sep 1;34(9):2340-2354. doi: 10.1093/molbev/msx170. PMID: 28541477 [Similar articles](#)

4. [Nuclear and mitochondrial RNA editing systems have opposite effects on protein diversity.](#) Sloan DB. Biol Lett. 2017 Aug;13(8). pii: 20170314. doi: 10.1098/rsbl.2017.0314. PMID: 28855414 [Similar articles](#)

5. [The Mitonuclear Dimension of Neanderthal and Denisovan Ancestry in Modern Human Genomes.](#) Sharbrough J, Havird JC, Noe GR, Warren JM, Sloan DB. Genome Biol Evol. 2017 Jun 1;9(6):1567-1581. doi: 10.1093/gbe/evx114. PMID: 28854627 [Free PMC Article](#) [Similar articles](#)

6. [The on-again, off-again relationship between mitochondrial genomes and species boundaries.](#) Sloan DB, Havird JC, Sharbrough J. Mol Ecol. 2017 Apr;26(8):2212-2236. doi: 10.1111/mec.13959. Epub 2017 Jan 27. Review. PMID: 27997046 [Similar articles](#)

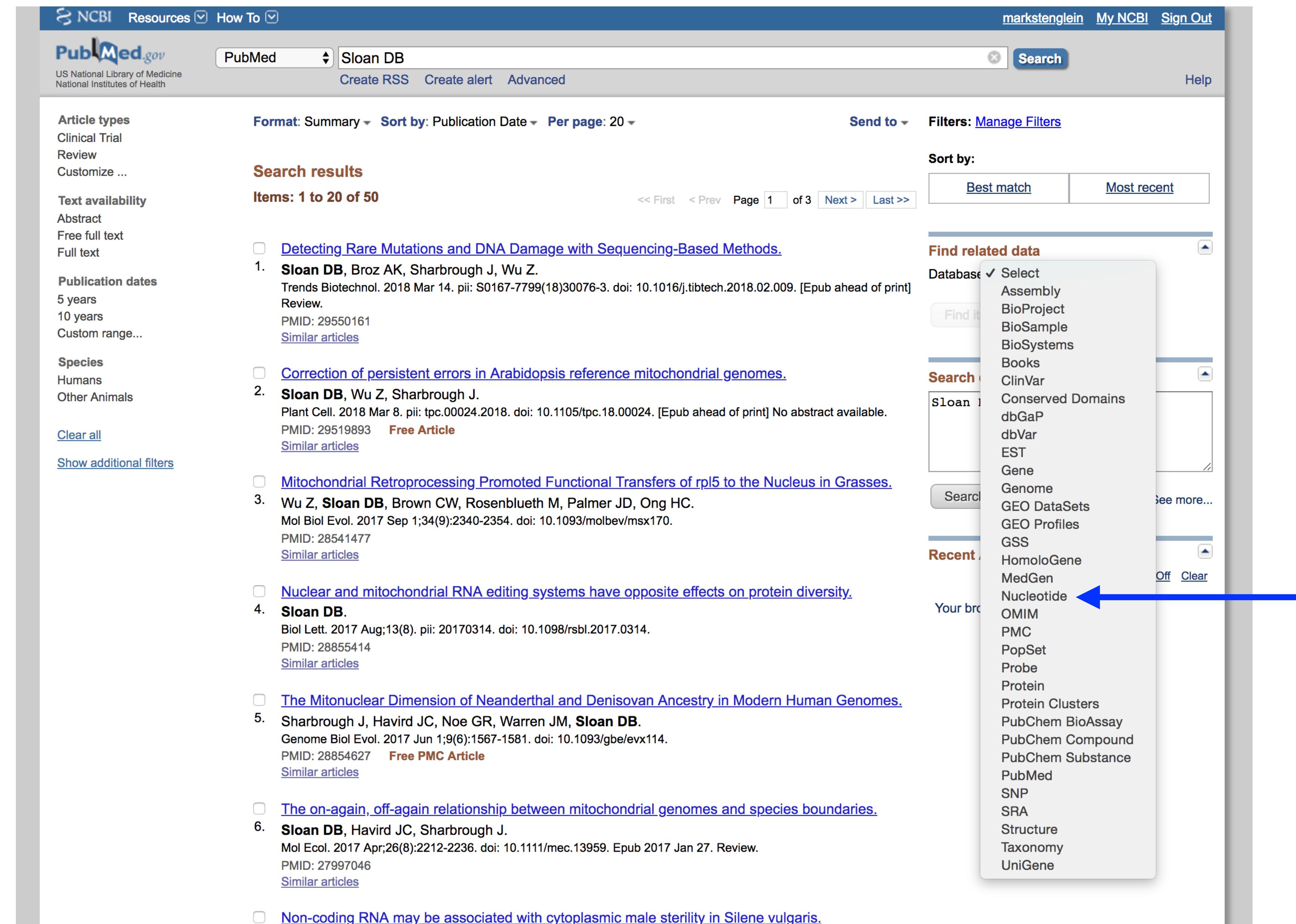
7. [Non-coding RNA may be associated with cytoplasmic male sterility in Silene vulgaris.](#)

Sort by: Best match Most recent

Find related data Database Select Assembly BioProject BioSample BioSystems Books ClinVar Conserved Domains dbGaP dbVar EST Gene Genome GEO DataSets GEO Profiles GSS HomoloGene MedGen Nucleotide OMIM PMC PopSet Probe Protein Protein Clusters PubChem BioAssay PubChem Compound PubChem Substance PubMed SNP SRA Structure Taxonomy UniGene

Recent Searches

Your browser



Nucleotide

Nucleotide

Advanced

Search

Help

Species

[Plants \(1,516\)](#)
[Fungi \(4\)](#)
[Bacteria \(19\)](#)
[Customize ...](#)

Molecule types

[genomic DNA/RNA \(1,417\)](#)
[mRNA \(116\)](#)
[rRNA \(6\)](#)
[Customize ...](#)

Source databases

[INSDC \(GenBank\) \(1,518\)](#)
[RefSeq \(20\)](#)
[Customize ...](#)

Genetic compartments

[Chloroplast \(209\)](#)
[Mitochondrion \(1,225\)](#)
[Plastid \(209\)](#)

Sequence length

[Custom range...](#)

Release date

[Custom range...](#)

Revision date

[Custom range...](#)[Clear all](#)[Show additional filters](#)

Summary ▾ 20 per page ▾ Sort by Default order ▾

Send to: ▾

Filter your results:

[All \(1539\)](#)[RefSeq \(20\)](#)[Manage Filters](#)

Items: 1 to 20 of 1539

<< First < Prev Page 1 of 77 Next > Last >>

- [Silene vulgaris isolate STK nad4L-atp4 intergenic spacer, complete sequence; mitochondrial](#)
1. 140 bp linear DNA
Accession: EF674110.1 GI: 157166813
[PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#) [PopSet](#)
- [Silene vulgaris isolate SAA nad4L-atp4 intergenic spacer, complete sequence; mitochondrial](#)
2. 140 bp linear DNA
Accession: EF674098.1 GI: 157166801
[PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#) [PopSet](#)
- [Silene vulgaris isolate SG nad4L-atp4 intergenic spacer, complete sequence; mitochondrial](#)
3. 140 bp linear DNA
Accession: EF674099.1 GI: 157166802
[PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#) [PopSet](#)
- [Silene vulgaris isolate VZA nad4L-atp4 intergenic spacer, complete sequence; mitochondrial](#)
4. 140 bp linear DNA
Accession: EF674100.1 GI: 157166803
[PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#) [PopSet](#)
- [Silene vulgaris isolate PIS nad4L-atp4 intergenic spacer, complete sequence; mitochondrial](#)
5. 140 bp linear DNA
Accession: EF674101.1 GI: 157166804
[PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#) [PopSet](#)

Results by taxon

Top Organisms [Tree]

[Silene vulgaris \(427\)](#)
[Silene latifolia \(320\)](#)
[Silene conica \(194\)](#)
[Silene noctiflora \(169\)](#)
[Silene paradoxa \(34\)](#)
[All other taxa \(395\)](#)[More...](#)

Find related data

Database: [Select](#)[Find items](#)

Recent activity

Your browsing activity is emp



Silene latifolia. image: sannse/Wikipedia

You could click on these sequences one at a time

NCBI Resources How To

Nucleotide Nucleotide Advanced Search Help

GenBank Send to: Change region shown

Customize view

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

Related information PubMed
Taxonomy
PopSet

Recent activity Turn Off Clear
Silene vulgaris isolate STK nad4L-atp4 intergenic spacer, complete sequenc Nucleotide See more...

Silene vulgaris isolate STK nad4L-atp4 intergenic spacer, complete sequence; mitochondrial

GenBank: EF674110.1

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

Go to:

LOCUS EF674110 140 bp DNA linear PLN 21-FEB-2008

DEFINITION Silene vulgaris isolate STK nad4L-atp4 intergenic spacer, complete sequence; mitochondrial.

ACCESSION EF674110

VERSION EF674110.1

KEYWORDS .

SOURCE mitochondrion Silene vulgaris (bladder campion)

ORGANISM [Silene vulgaris](#)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae;
Pentapetalae; Caryophyllales; Caryophyllaceae; Sileneae; Silene.

REFERENCE 1 (bases 1 to 140)

AUTHORS Sloan,D.B., Barr,C.M., Olson,M.S., Keller,S.R. and Taylor,D.R.

TITLE Evolutionary rate variation at multiple levels of biological organization in plant mitochondrial DNA

JOURNAL Mol. Biol. Evol. 25 (2), 243-246 (2008)

PUBMED [18056075](#)

REFERENCE 2 (bases 1 to 140)

AUTHORS Sloan,D.B., Barr,C.M., Olson,M.S. and Taylor,D.R.

TITLE Direct Submission

JOURNAL Submitted (12-JUN-2007) Biology Department, University of Virginia,
Gilmer Hall, Charlottesville, VA 22904, USA

FEATURES Location/Qualifiers

source 1..140
/organism="Silene vulgaris"
/organelle="mitochondrion"
/mol_type="genomic DNA"
/isolate="STK"
/db_xref="taxon:[42043](#)"

misc_feature 1..140
/note="nad4L-atp4 intergenic spacer"

ORIGIN

1 tcgttctttt cttttttttt ttatttaaat ttacaaaat ggcagggtca gggctttct
61 cgctggcgaa ggcacccgaa ttcaactttt cctaaagaaac atcccgttca gtttgtaaaa
121 gataaaagata agctttataa
//

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

NCBI Resources ▾ How To ▾

Nucleotide Nucleotide Advanced Search Help

Species Summary ▾ 20 per page ▾ Sort by Default order ▾

Plants (1,516)
Fungi (4)
Bacteria (19)
Customize ...

Molecule types genomic DNA/RNA (1,417)
mRNA (116)
rRNA (6)
Customize ...

Source databases INSDC (GenBank) (1,518)
RefSeq (20)
Customize ...

Genetic compartments Chloroplast (209)
Mitochondrion (1,225)
Plastid (209)

Sequence length Custom range...

Release date Custom range...

Revision date Custom range...

[Clear all](#)
[Show additional filters](#)

Items: 1 to 20 of 1539 << First < Prev Page 1 of 80 > Next >> Last

[Silene vulgaris isolate STK nad4L-atp4 intergenic spacer, complete sequence; mitochondrial](#)
1. 140 bp linear DNA
Accession: EF674110.1 GI: 157166813
[PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#) [PopSet](#)

[Silene vulgaris isolate SAA nad4L-atp4 intergenic spacer, complete sequence; mitochondrial](#)
2. 140 bp linear DNA
Accession: EF674098.1 GI: 157166801
[PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#) [PopSet](#)

[Silene vulgaris isolate SG nad4L-atp4 intergenic spacer, complete sequence; mitochondrial](#)
3. 140 bp linear DNA
Accession: EF674099.1 GI: 157166802
[PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#) [PopSet](#)

[Silene vulgaris isolate VZA nad4L-atp4 intergenic spacer, complete sequence; mitochondrial](#)
4. 140 bp linear DNA
Accession: EF674100.1 GI: 157166803
[PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#) [PopSet](#)

[Silene vulgaris isolate PIS nad4L-atp4 intergenic spacer, complete sequence; mitochondrial](#)
5. 140 bp linear DNA

Send to: ▾ Filter your results:

Complete Record
 Coding Sequences
 Gene Features

Choose Destination

File Clipboard
 Collections

Download 1539 items.

Format

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

Manage Filters

Find related data

Database: Select

Find items

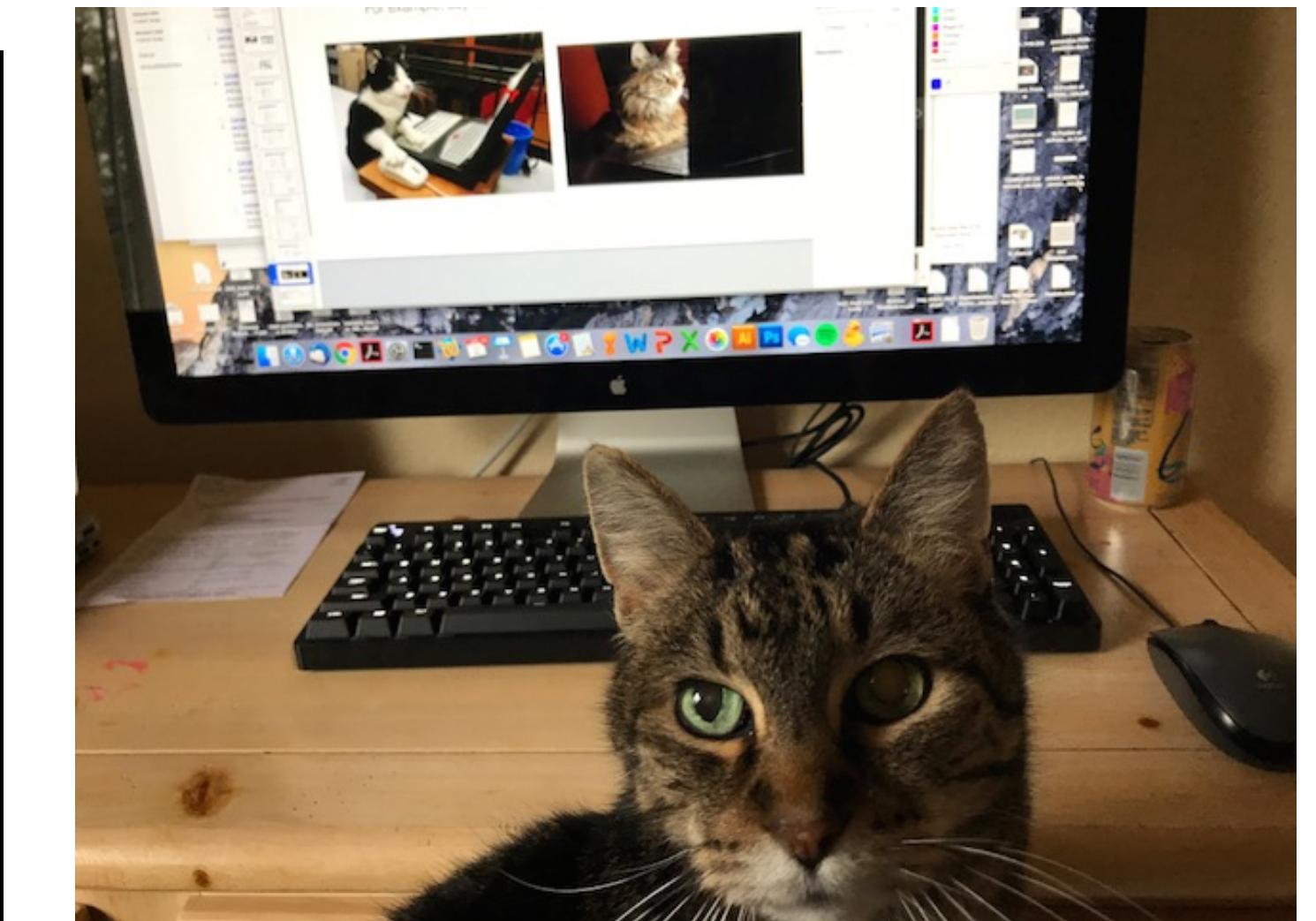
Recent activity

Turn Off Clear

Your browsing activity is empty.

There are often many paths to the same data

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



Kirby

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...

A better way to get the cat genome would be via the Genome database

← → C 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: 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...](#)

There are actually 2 cat genome assemblies in NCBI

← → C 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: ID: 78

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

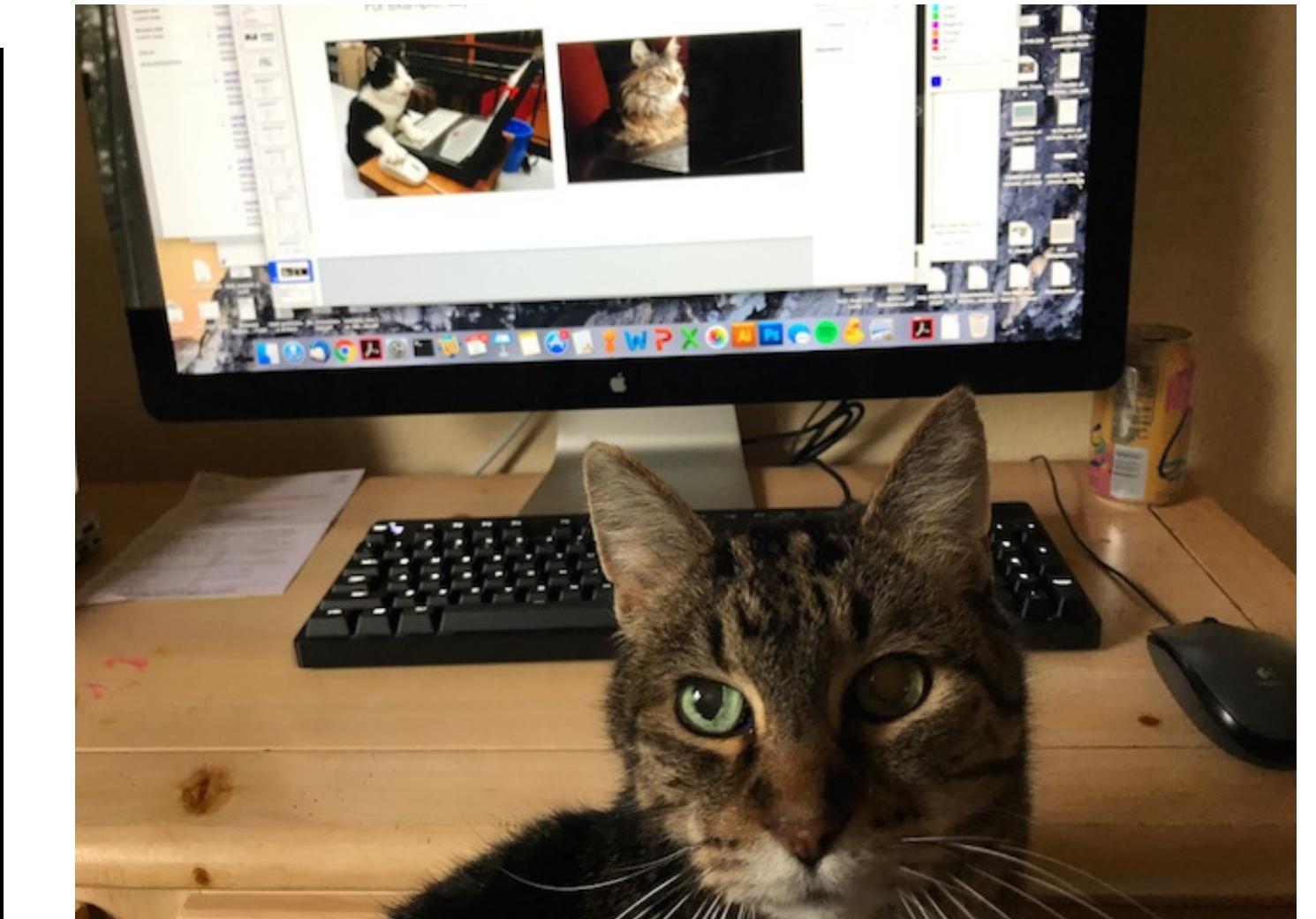
 **Felis catus (domestic cat)**
domestic cat

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

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)

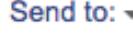
Comment
This assembly adds 0.9X coverage to the 1.9X coverage of the previous cat assembly (WGS project AANG00000000 of the Abyssinian cat "Cinnamon"). The 0.9X coverage comes from six other breeds and one wild cat from paired-end fosmid ABI 3730 ... [more](#)

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: [SAMN02953640](#)
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: [AANG03](#)
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	2,641,342,258
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	367,672
Contig N50	45,189
Contig L50	16,252
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)
 - [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](#)
 - [environmental samples](#)
- [Otocolobus](#) [1](#)

You can go up the taxonomic tree in the Taxonomy db

NCBI Resources How To

Genome Genome txid32536[Organism:exp] | Create alert Limits Advanced

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]; [Felinea](#)[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...](#)

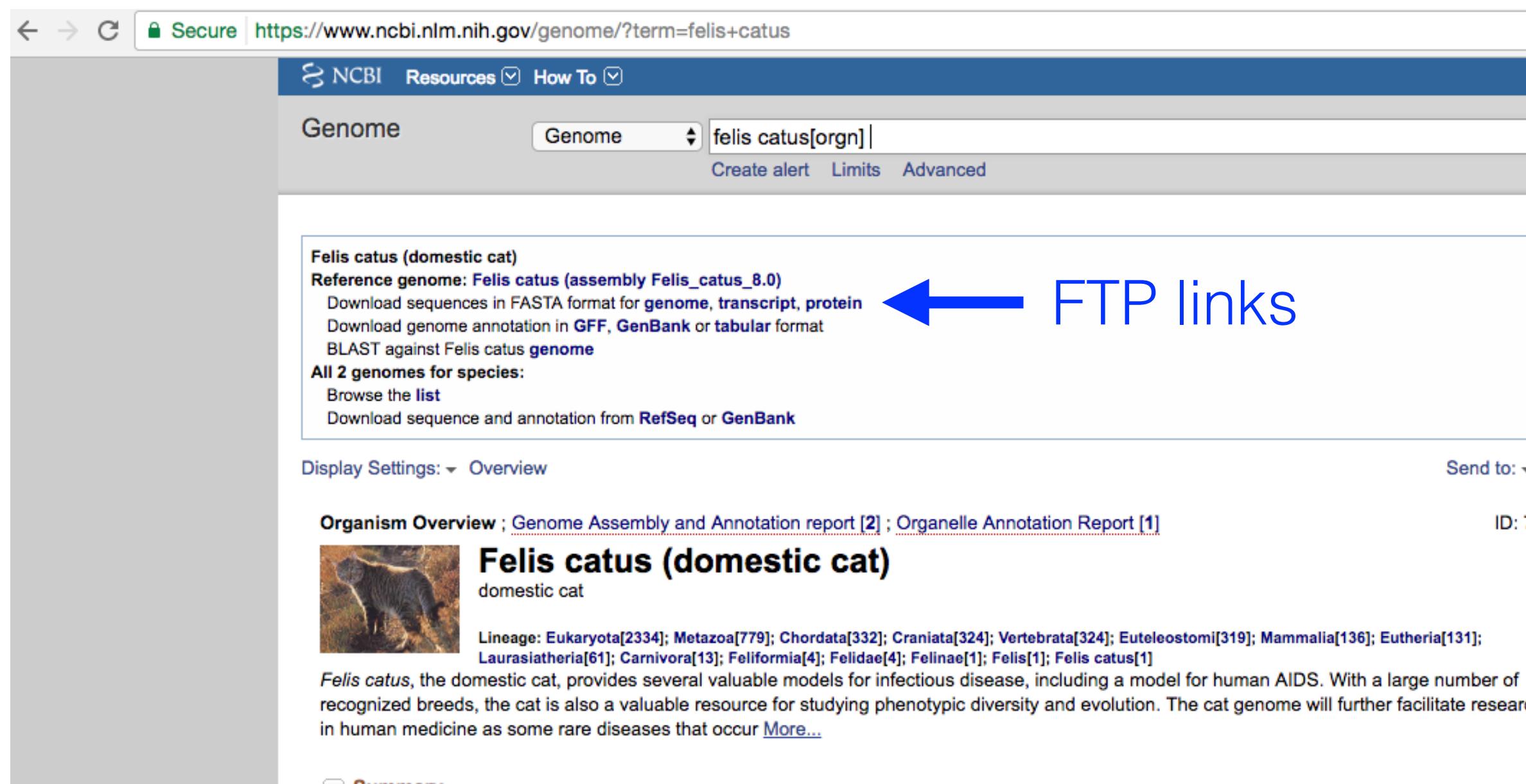
← → C i <ftp://ftp.ncbi.nlm.nih.gov/genomes/>

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:

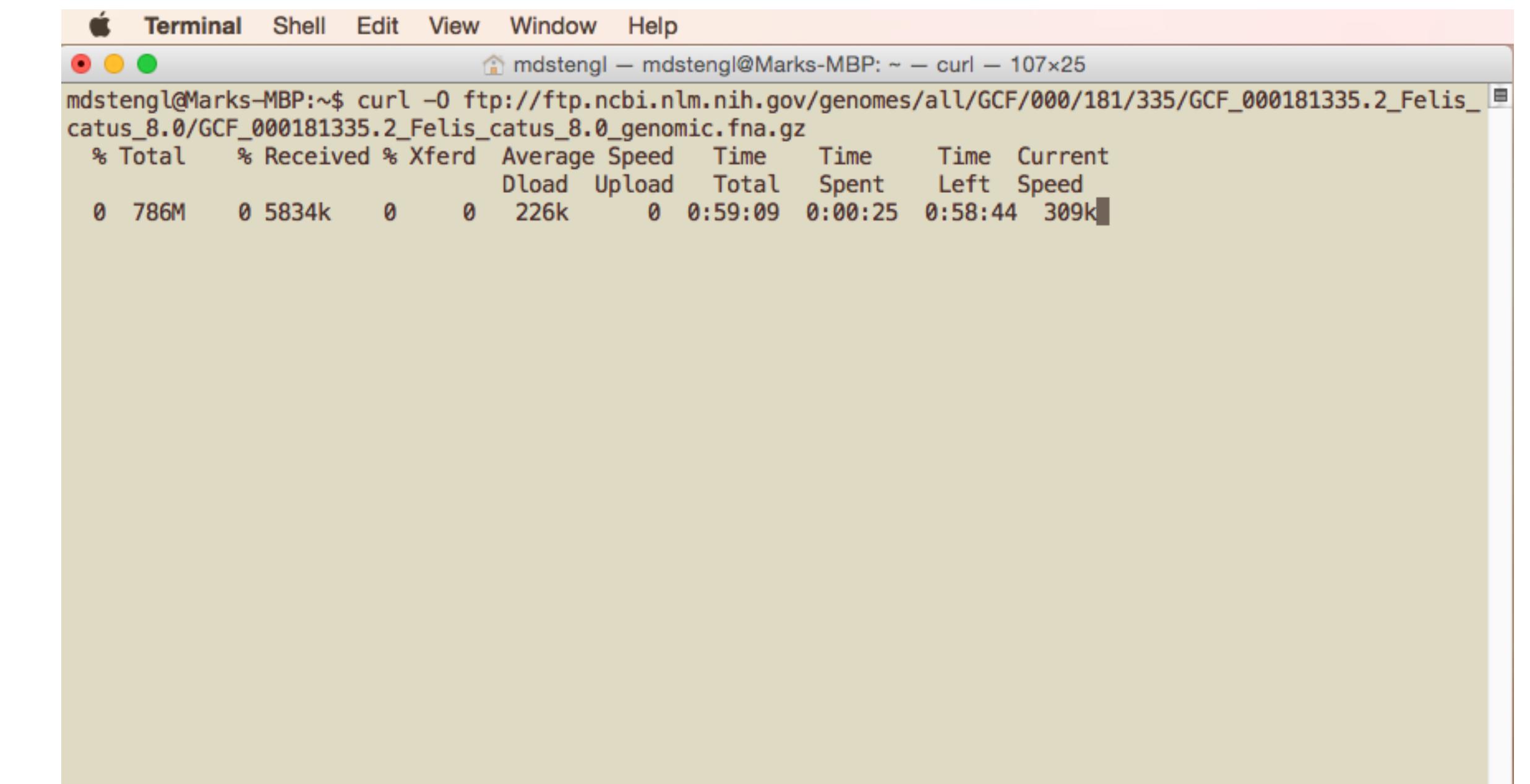
Organism Overview ; Genome Assembly and Annotation report [2] ; Organelle Annotation Report [1]

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



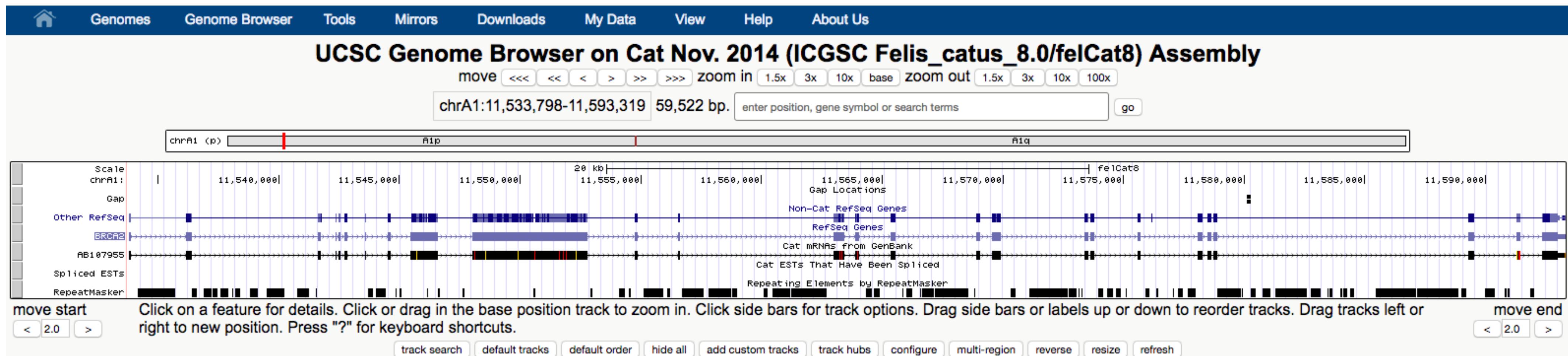
Screenshot of the Cyberduck application interface. The title bar shows "ftp.ncbi.nlm.nih.gov - FTP" and "Unregistered". The main window displays a file list with columns for "Filename", "Size", and "Modified". The list includes various directories and files such as "genomes", "bioproject", "biosample", "snp", "repository", "pubchem", "pathogen", "hmm", "geo", "pubmed", "refseq", "sra", "ncbi-asn1", "genbank", "eqtl", "dbgap", "pub", "sequin", "variation", and "nist-immsa". The total count at the bottom is "46 Files".

Filename	Size	Modified
genomes	32.8 KB	Today, 9:23 AM
bioproject	4.1 KB	Today, 9:10 AM
biosample	4.1 KB	Today, 2:39 AM
snp	4.1 KB	Yesterday, 8:48 PM
repository	4.1 KB	Yesterday, 8:48 PM
pubchem	4.1 KB	Yesterday, 8:48 PM
pathogen	4.1 KB	Yesterday, 8:48 PM
hmm	4.1 KB	Yesterday, 1:20 PM
geo	0 B	Yesterday, 6:22 AM
pubmed	4.1 KB	5/24/17, 2:05 AM
refseq	4.1 KB	5/15/17, 11:38 AM
sra	4.1 KB	5/4/17, 11:38 AM
ncbi-asn1	73.7 KB	4/19/17, 9:20 PM
genbank	94.2 KB	4/19/17, 9:07 PM
eqtl	4.1 KB	4/15/17, 4:14 PM
dbgap	4.1 KB	4/3/17, 1:34 PM
pub	8.2 KB	3/20/17, 10:58 AM
sequin	4.1 KB	2/22/17, 6:50 AM
variation	4.1 KB	8/9/16, 8:36 AM
nist-immsa	4.1 KB	6/30/16, 12:19 PM

<ftp://ftp.ncbi.nlm.nih.gov/>



Genome browsers, like Ensembl and UCSC, offer additional functionality



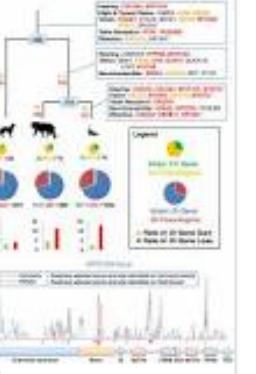
Finally, there's nothing wrong with using Google

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 ...](http://www.pnas.org/content/111/48/17230.abstract)
www.pnas.org/content/111/48/17230.abstract

[About this result](#) [Feedback](#)

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.

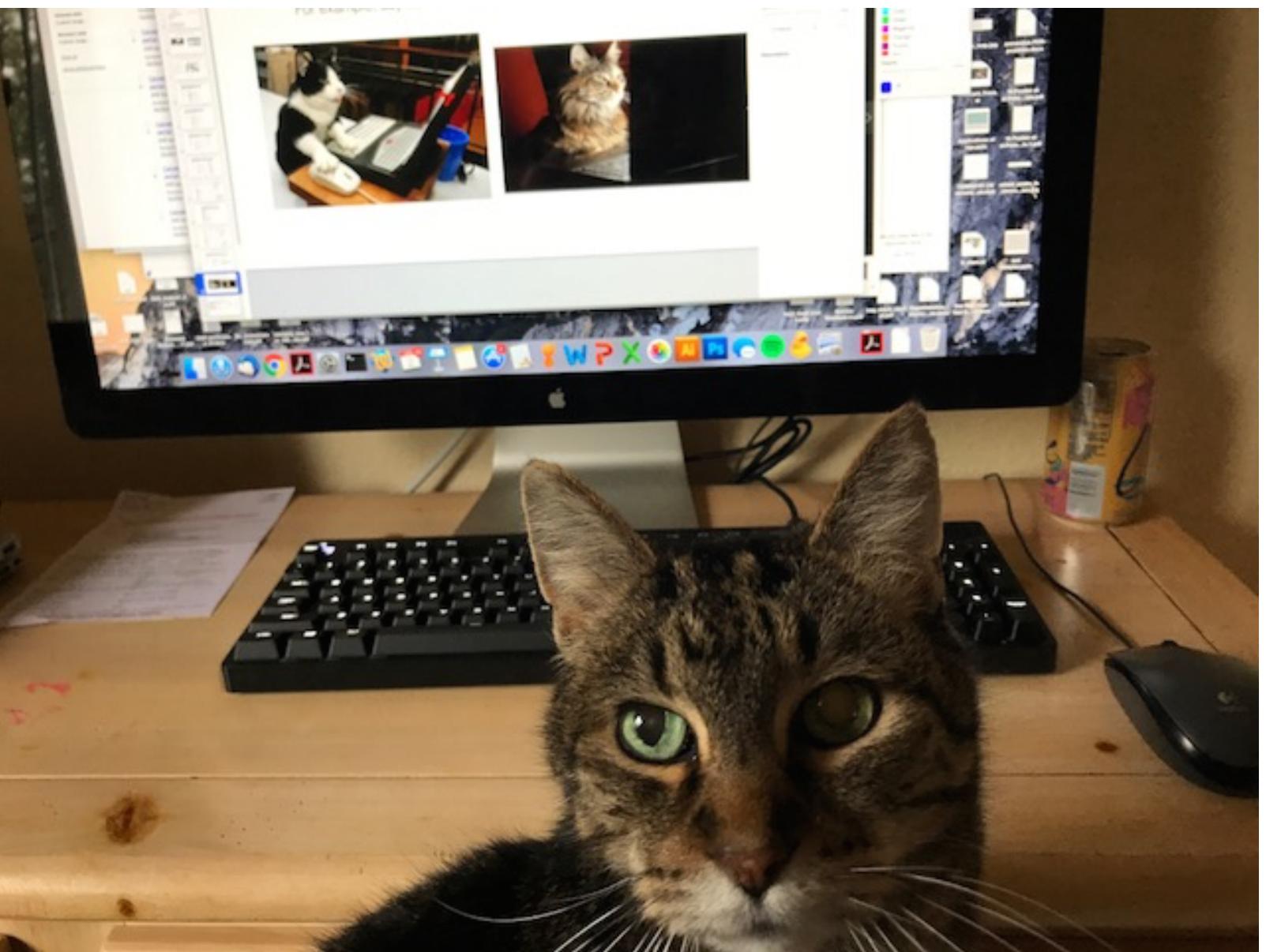
[Comparative analysis of the domestic cat genome reveals genetic ...](http://www.pnas.org/content/111/48/17230.abstract)
www.pnas.org/content/111/48/17230.abstract ▾
by MJ Montague - 2014 - Cited by 67 - Related articles
Dec 2, 2014 - Specifically, the signatures of selection in the domestic **cat genome** 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 ...](http://www.nature.com/news/i-can-haz-genomes-cats-claw-their-way-into-genetics-1.16708)
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