

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

Mark Stenglein

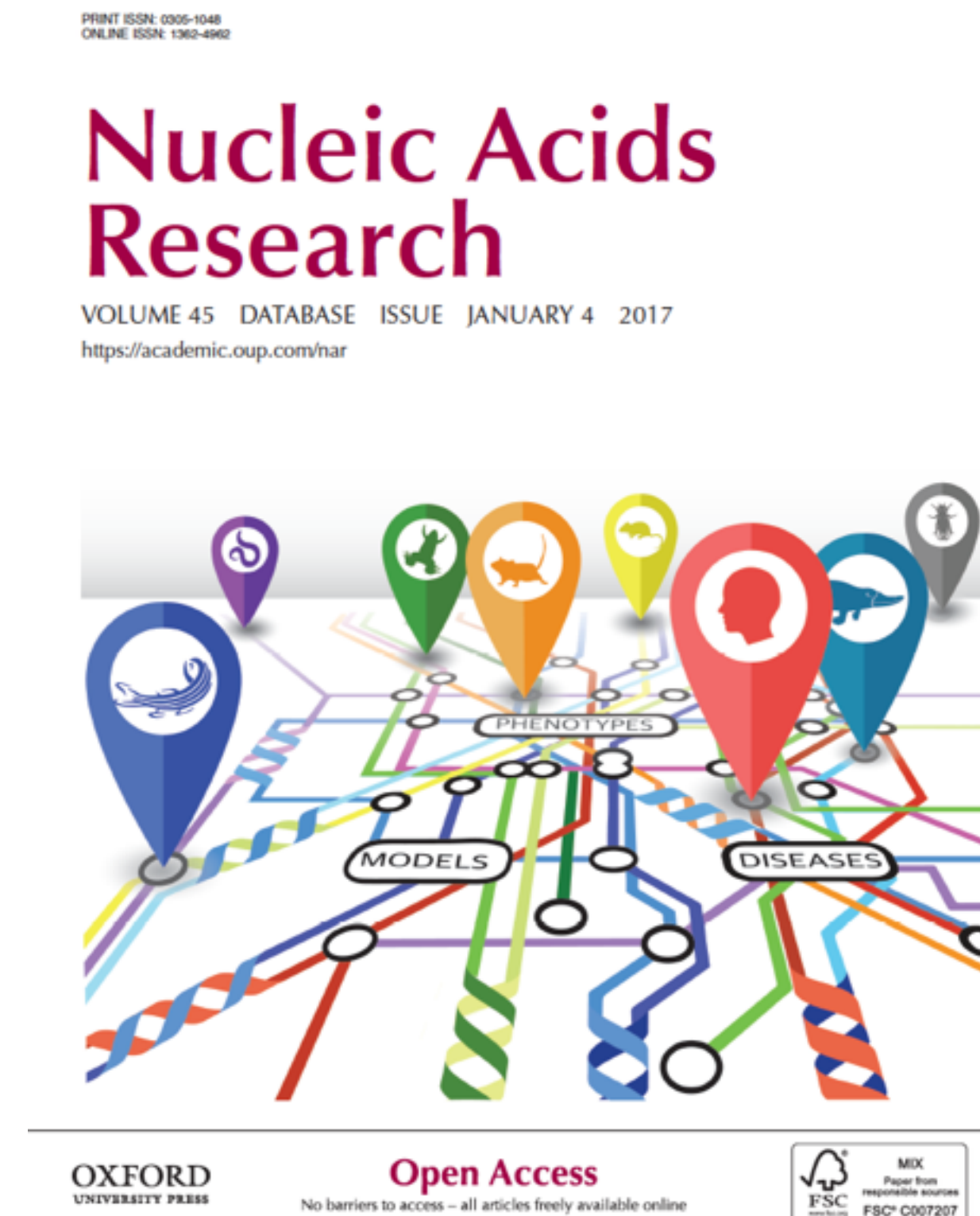


Computational Biology and
Genomics Workshop

Todos Santos Center
April 9-13, 2018

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

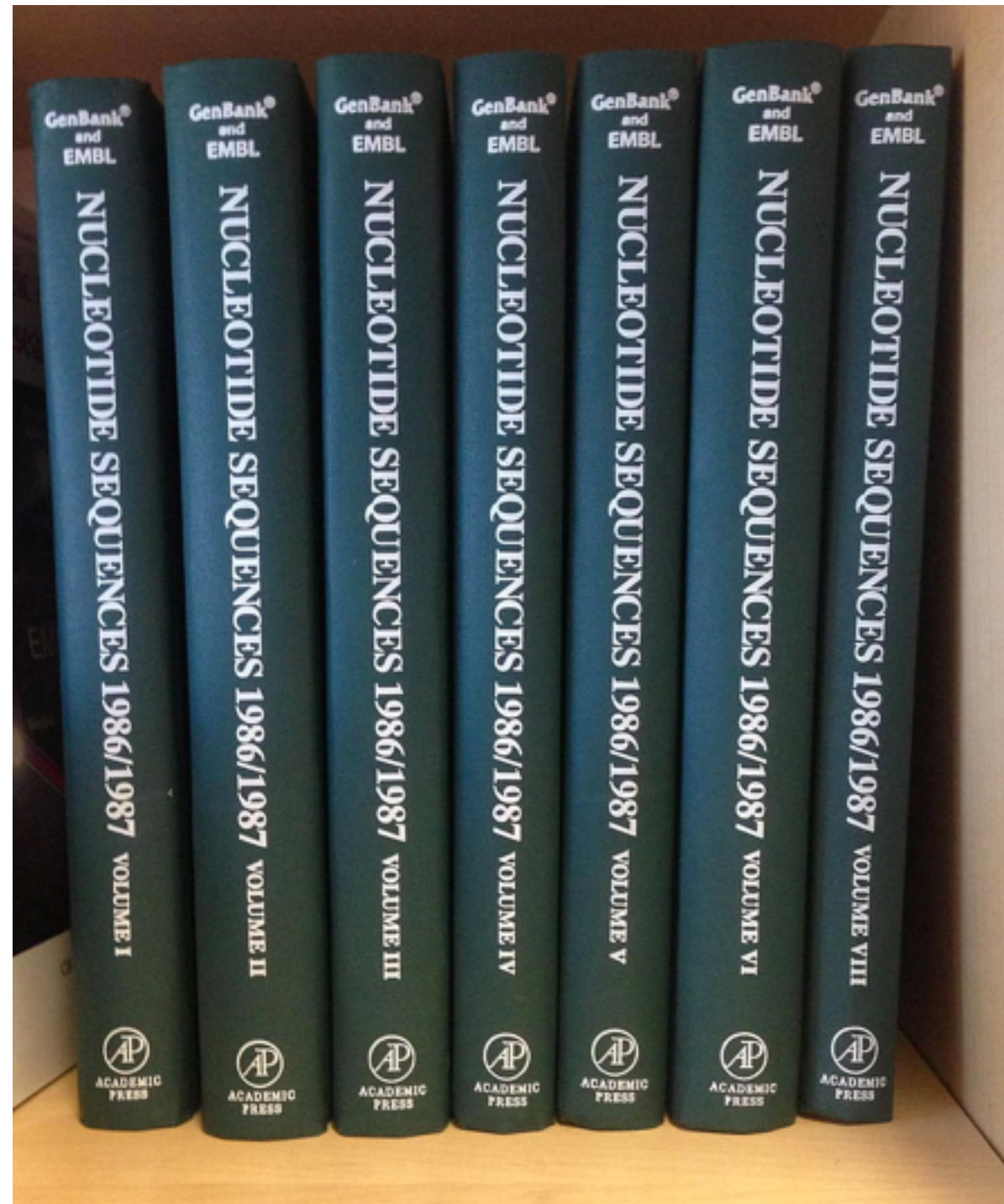
The annual Database and Web Server NAR issue is a good resource



<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

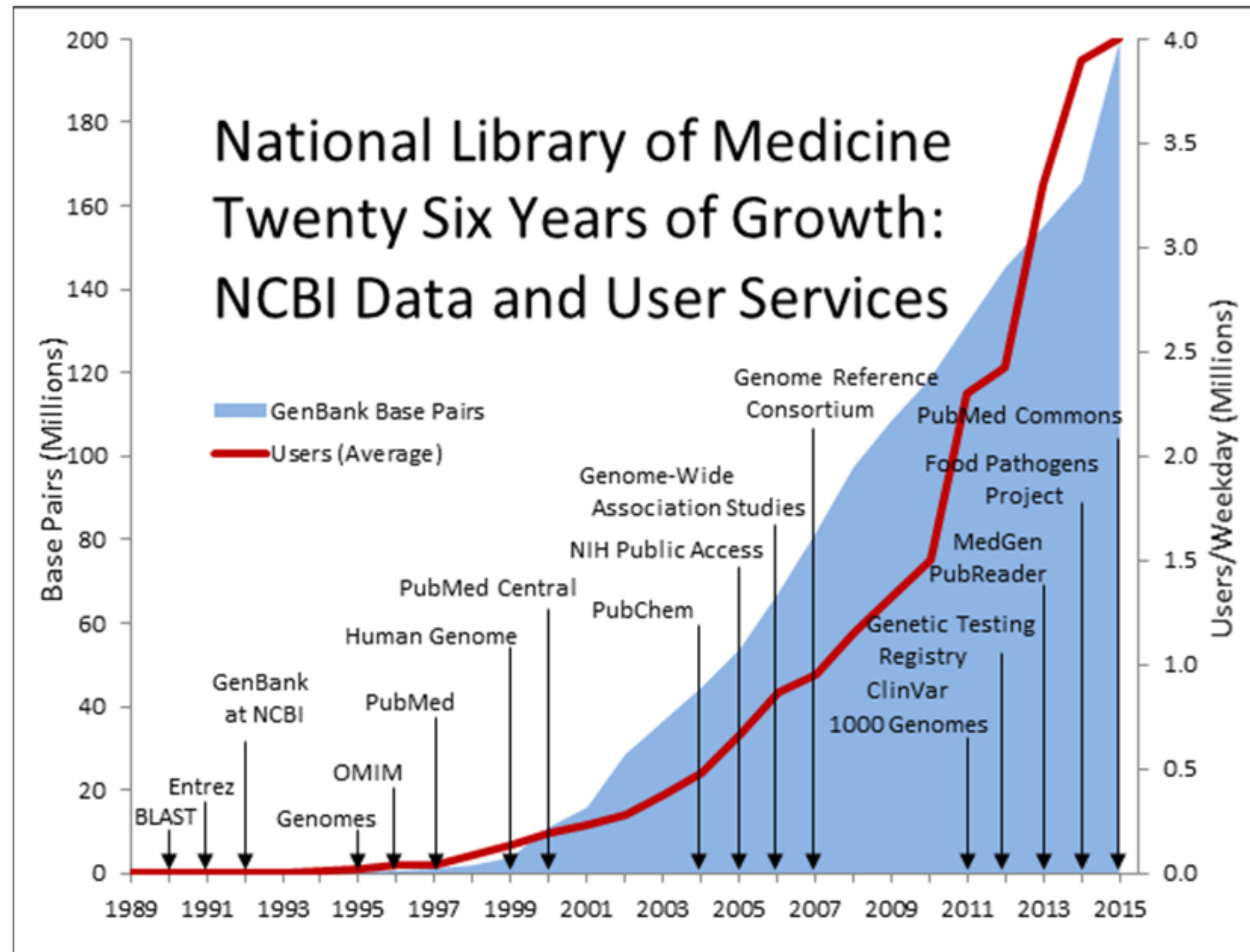


>200,000,000 sequences

```
NUCLEOTIDE SEQUENCES 1984
SITE:
key      site span  description      key      site span  description
refnumbr 21 1 numbered 1 in [1]      pept/pept 195 0 chymoa propept end/ mature pept
->pept    21 1 chymoa prepropept cds start
pept/pept 69 0 chymoa prepropept end/ propept start pept<- 1166 1 chymoa mature pept cds end
ORIGIN:  20 bases upstream from codon 1
SEQUENCE: 1275 bp 293 a 391 c 336 g 255 t
1  cggctggacc  cagatccaag  atgaggtgtc  tctgtgtct  acttctctc  ttcgtctct  cccagggggc  tgagatcacc  aggatccctc  tgtacaagg
101 caagtctctg  aggaaggcgc  tgaaggagca  tgggtctctg  gaggacttcc  tgcagaacca  gtagtatgpc  atcagcagca  agtactctcg  cttcggggag
201 gtggccagrg  tggccctgac  caactactct  gatagtcagt  actttgggaa  gatctacttc  gggaccctgc  cccaggagtt  caccgtctgt  tttagactgt
301 gctcctctga  cttctgggta  cctctctatc  actgcaagag  caatgctctc  aaaaaccacc  agcgtctcga  cccgagaag  tigtccact  tccagaacct
401 gggcaagccc  ctgtctatcc  actacgggac  aggcagctg  caggcctctc  tgggctatga  caccgtctct  gttcccaaca  ttctcgacct  ccagcagaca
501 gtaggcctga  gcaaccagga  gcccggggac  gtcttcacct  atpccgaatt  cgaagggatc  ctggggatgg  cctacccttc  gctcgcctca  pagtactcga
601 tacccctgtt  tgaacaactg  atgaacagcc  acctggtgpc  ccaapacttg  ttctcggttt  acatggacag  gaatggccag  gagagcatgc  tcacgtctgg
701 ggcactcgac  ccgtctact  acacagggtc  cctgacttgg  gtgcccttga  cagtgcagca  gtactgtcag  ttactgtgg  acagtgtcac  catcagtggt
801 gtggttgggg  cctgtgaggg  tggctgtcag  gccatcttgg  acagggccac  ctccaagctg  gtccggccca  gcagcagcat  cctcaacatc  cagcagggca
901 ttggagccac  acagaaccag  tacgatgagt  ttgacatcga  ctgpcacaac  ctgagctaca  tgcacctgtg  ggtcttttgg  atcaatggca  aaatgtacc
1001 actgaccctc  tcgcctata  ccagccagga  ccagggcttc  tgtaccagt  gcttccagag  tgaaaatcat  tcccagaat  gpatctctgg  gpatgttttc
1101 atccgagagt  attacagct  ctttgacagg  gccaaacc  tctggggct  ggcacaagcc  atctgatcac  atcgtgacc  aagaacctca  ctgtccacc
1201 acacctgcac  acacacatgc  acacatgtac  atggcactg  tgcacacaca  cagatgaggt  ttccagacc  aagct
```

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



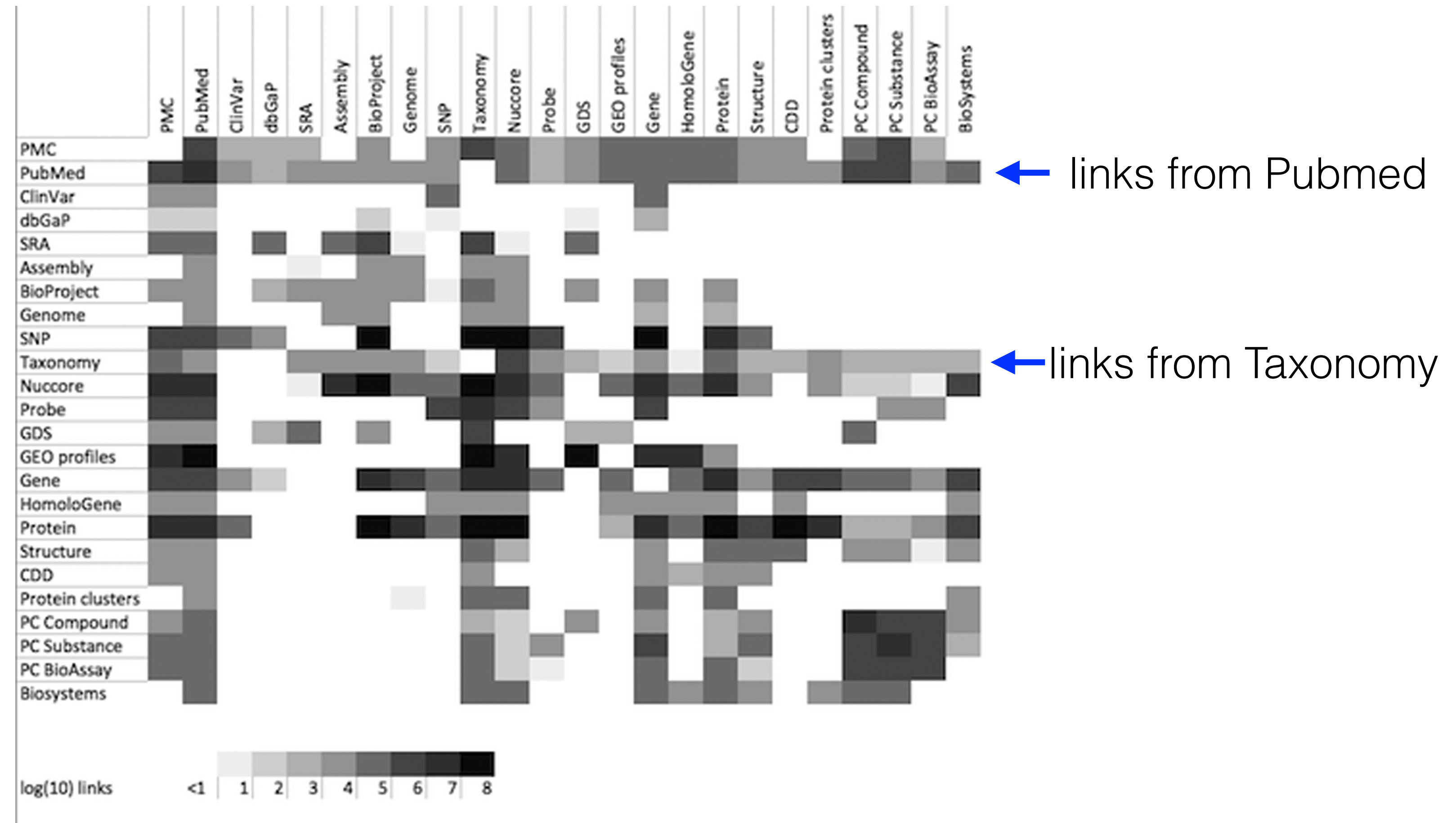
Categories of NCBI databases

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

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

The screenshot shows the PubMed search results for 'Sloan DB'. The search results are sorted by 'Publication Date' and show 6 items. A blue arrow points to the 'Find related data' dropdown menu on the right side of the page, which is currently set to 'Select'.

NCBI Resources How To markstenglein My NCBI Sign Out

PubMed Sloan DB Search

US National Library of Medicine National Institutes of Health Create RSS Create alert Advanced Help

Article types: 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

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 >>

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.

- [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](#)

Get nucleotide sequences associated with Dan's publications

The screenshot shows the PubMed website interface. At the top, there is a navigation bar with 'NCBI Resources' and 'How To' menus, and a user profile for 'markstenglein'. Below this is the 'PubMed.gov' logo and a search bar containing 'Sloan DB'. The search results are displayed in a list format, with the first item selected. A dropdown menu is open over the 'Find related data' section, listing various databases. A blue arrow points to the 'Nucleotide' option in the dropdown menu.

NCBI Resources How To markstenglein My NCBI Sign Out

PubMed Sloan DB Search

US National Library of Medicine National Institutes of Health

Article types: 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

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 >>

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

Your browser

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](#)

Non-coding RNA may be associated with cytoplasmic male sterility in Silene vulgaris.

Nucleotide

Nucleotide

Search

Advanced

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:

[Find items](#)

Recent activity

Your browsing activity is empty



Silene latifolia. image: sannse/Wikipedia

You could click on these sequences one at a time

NCBI Resources How To markstenglein My NCBI Sign Out

Nucleotide Nucleotide Search Help

Advanced

GenBank Send to: Change region shown

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 tcgttctctt ctttctttt ttatttaaatt tacaaagtt ggcagggtca ggcctttct
61 cgctgggcga ggcaccgca ttcgaacttt cctaagaac atcccgttca gttgctgaaa
121 gataaagata agctttataa
//

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

Related information
PubMed
Taxonomy
PopSet

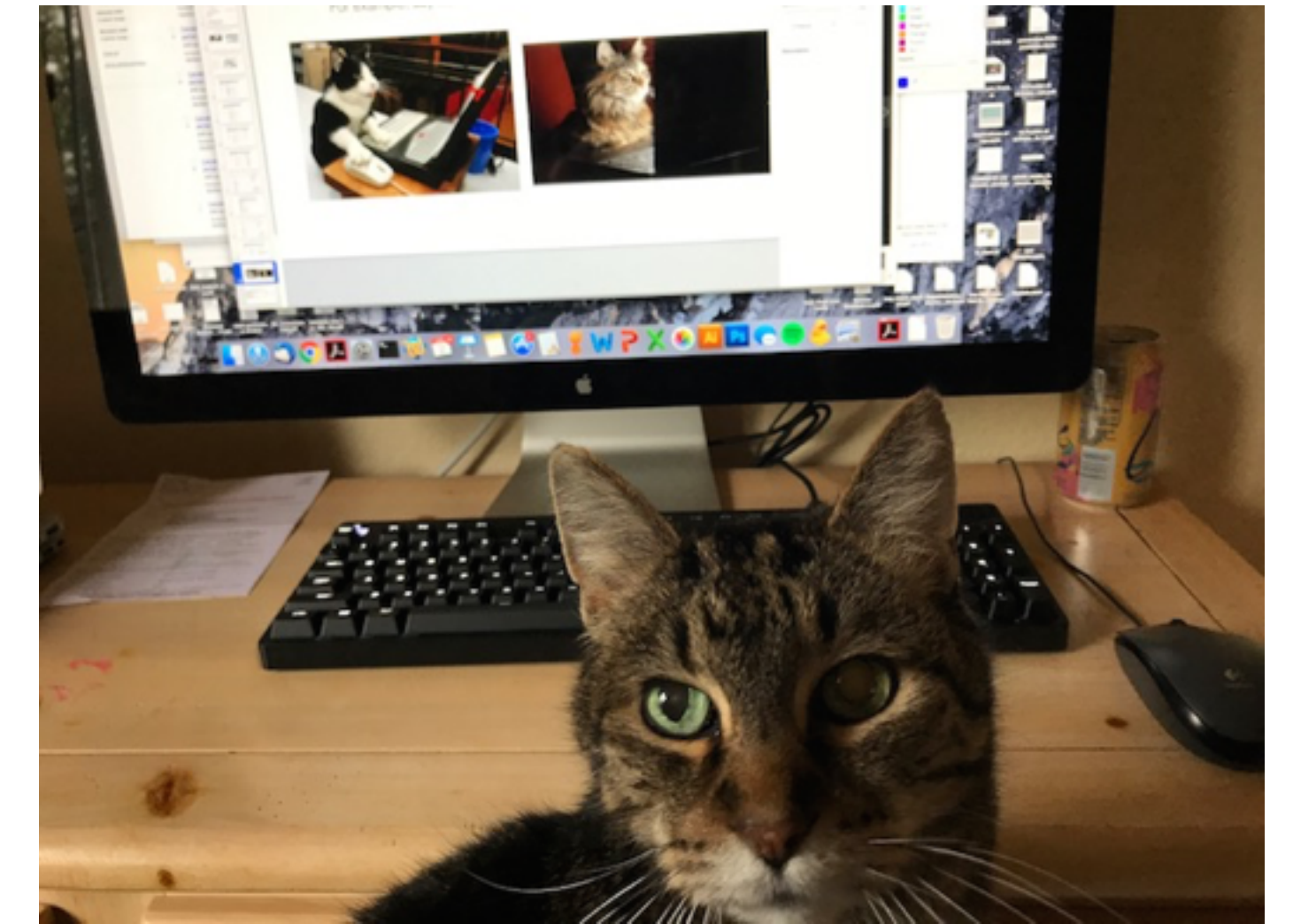
Recent activity
Turn Off Clear
Silene vulgaris isolate STK nad4L-atp4 intergenic spacer, complete sequenc Nucleotide
See more...

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

The screenshot shows the NCBI Nucleotide search interface. The top navigation bar includes the NCBI logo, 'Resources', 'How To', and user links for 'markstenglein', 'My NCBI', and 'Sign Out'. The main search area has a 'Nucleotide' dropdown, a search input field, and a 'Search' button. Below the search bar, there are filters for 'Species' (Plants, Fungi, Bacteria), 'Molecule types' (genomic DNA/RNA, mRNA, rRNA), and 'Source databases' (INSDC, RefSeq). The search results are displayed in a table with columns for 'Items: 1 to 20 of 1539', 'Summary', '20 per page', and 'Sort by Default order'. The first five items are listed, each with a checkbox, a title, accession number, GI number, and links to 'PubMed', 'Taxonomy', 'GenBank', 'FASTA', 'Graphics', and 'PopSet'. A 'Send to' dropdown menu is open, showing options for 'Complete Record', 'Coding Sequences', 'Gene Features', 'File', 'Clipboard', and 'Collections'. A sub-menu for 'Choose Destination' is also open, showing options for 'Summary', 'GenBank', 'GenBank (full)', 'FASTA', 'ASN.1', 'XML', 'INSDSeq XML', 'TinySeq XML', 'Feature Table', 'Accession List', 'GI List', and 'GFF3'. The 'Send to' menu is highlighted with a blue arrow. The 'Filter your results:' section is also visible, with a 'Manage Filters' link. The 'Find related data' section has a 'Database:' dropdown and a 'Find items' button. The 'Recent activity' section shows 'Your browsing activity is empty.' with 'Turn Off' and 'Clear' links.

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/nucleotide/?term=Felis+catus>

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide Felis catus Search

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 ...

Molecule types
genomic DNA/RNA (319,957)
mRNA (63,505)
rRNA (13)
Customize ...

Source databases
INSDC (GenBank) (37,020)
RefSeq (366,045)
Customize ...

Genetic compartments
Mitochondrion (860)
Plasmid (2)
Plastid (11)

Sequence length
Custom range...

Release date
Custom range...

Revision date

Items: 1 to 20 of 403089

<< First < Prev Page 1 of 20155 Next > Last >>

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 shotgun sequence](#)
2. 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 shotgun sequence](#)
3. 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 shotgun sequence](#)
4. 140,925,898 bp linear DNA
Accession: NC_018725.2 GI: 753572100
[GenBank](#) [FASTA](#) [Graphics](#)

Results by taxon

Top Organisms [Tree]

- [Felis catus \(355407\)](#)
- [Hammondia 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 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...](#)

Summary

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**
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

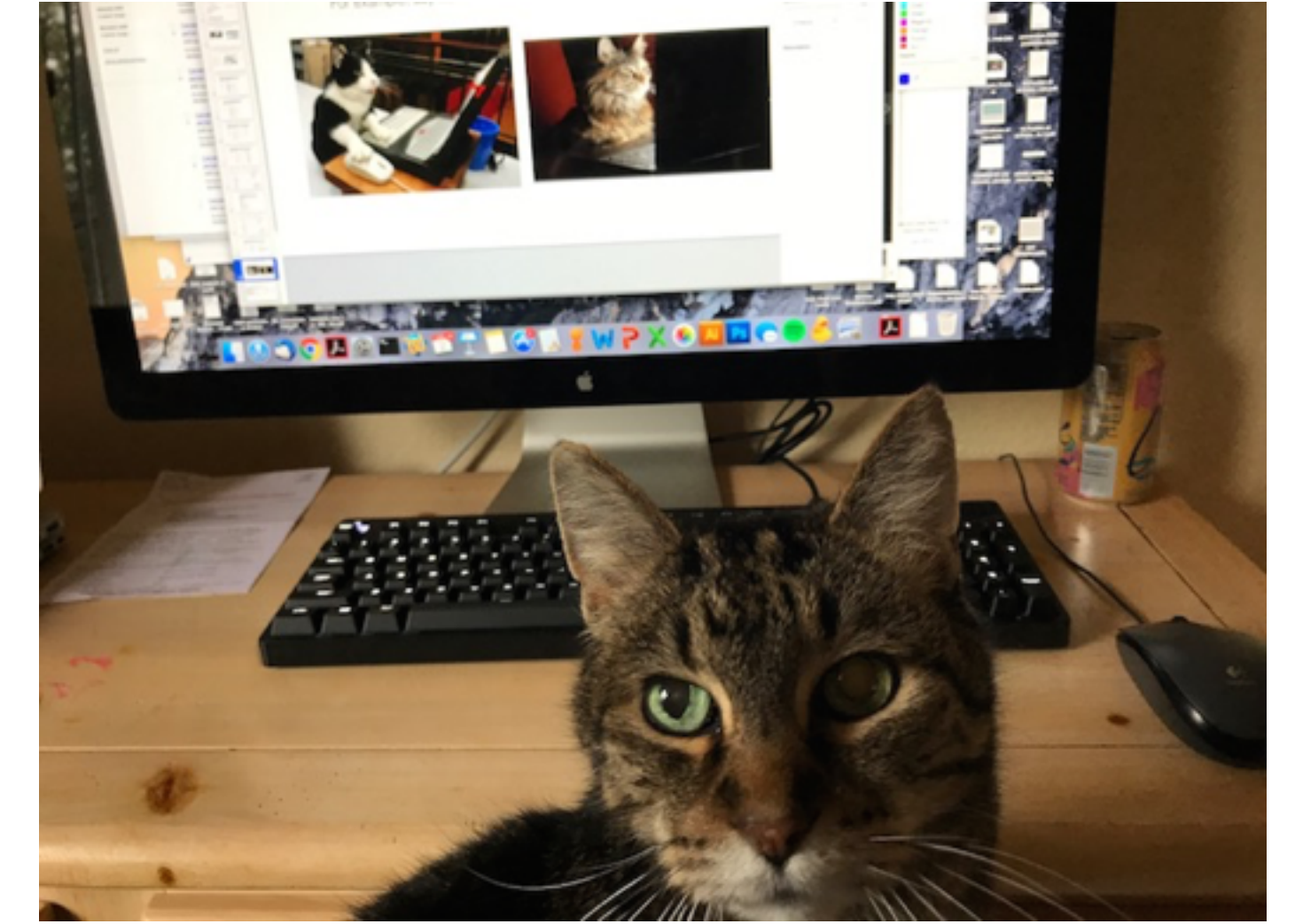
Lineage: [Eukaryota\[2334\]](#); [Metazoa\[779\]](#); [Chordata\[332\]](#); [Craniata\[324\]](#); [Vertebrata\[324\]](#); [Euteleostomi\[319\]](#); [Mammalia\[136\]](#); [Eutheria\[131\]](#); [Laurasiatheria\[61\]](#); [Carnivora\[13\]](#); [Feliformia\[4\]](#); [Felidae\[4\]](#); [Felinae\[1\]](#);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

Summary

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 Send to: ▾

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

[See Genome Information for Felis catus](#)

There are 2 assemblies for this organism
[See more](#)

NCBI Resources How To

Assembly Assembly | Advanced Browse by organism

Full Report Send to: ▾

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

[See Genome Information for Felis catus](#)

There are 2 assemblies for this organism
[See more](#)

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

Entrez PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books

Search for as lock

Display levels using filter:

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

Entrez PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books

Search for as lock

Display levels using filter:

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

The screenshot shows the NCBI Taxonomy Browser interface. At the top, there are logos for Entrez, PubMed, Nucleotide, Protein, and Genome. Below these is a search bar with the text "Search for" and a dropdown menu set to "AS complete name". There are also checkboxes for "lock" and a "Go" button. Below the search bar, there are several filters and options, including "Display 3 levels using filter: none". A blue arrow points to the "Display" dropdown. Below the filters, there is a list of checkboxes for various data types, including "Genome", "HomoloGene", "Viral Host", "Bio Project", "Bio Sample", "SNP", "Probe", and "Bio Sample". A blue arrow points to the "Genome" checkbox. Below the filters, there is a "Lineage" section with a list of taxonomic levels: root; cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Amniota; Mammalia; Theria; Eutheria; Boreoeutheria; Laurasiatheria; Carnivora; Feliformia. Below the lineage, there is a list of taxonomic entries for the Felidae family. A blue arrow points to the "Acinonyx jubatus (cheetah)" entry.

Search for lock

Display levels using filter:

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](#)

- o [Felidae](#) (cat family) **36** *Click on organism name to get more information.*
 - o [Acinonychinae](#) **1**
 - o [Acinonyx](#) **1**
 - [Acinonyx jubatus](#) (cheetah) **1**
 - o [Felinae](#) **28**
 - o [Caracal](#) **1**
 - [Caracal caracal](#) **1**
 - o [Catopuma](#) **2**
 - [Catopuma badia](#) (bay cat) **1**
 - [Catopuma temminckii](#) (Asiatic golden cat) **1**
 - o [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](#)
 - o [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**
 - o [Leptailurus](#) **1**
 - [Leptailurus serval](#) (serval) **1**
 - [Leptailurus serval x Caracal caracal](#)
 - o [Lynx](#) **4**
 - [Lynx canadensis](#) (Canada lynx) **1**
 - [Lynx lynx](#) (Eurasian lynx) **1**
 - [Lynx pardinus](#) (Spanish lynx) **1**
 - [Lynx rufus](#) (bobcat) **1**
 - [environmental samples](#)
 - o [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 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** ← **FTP links**
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...](#)

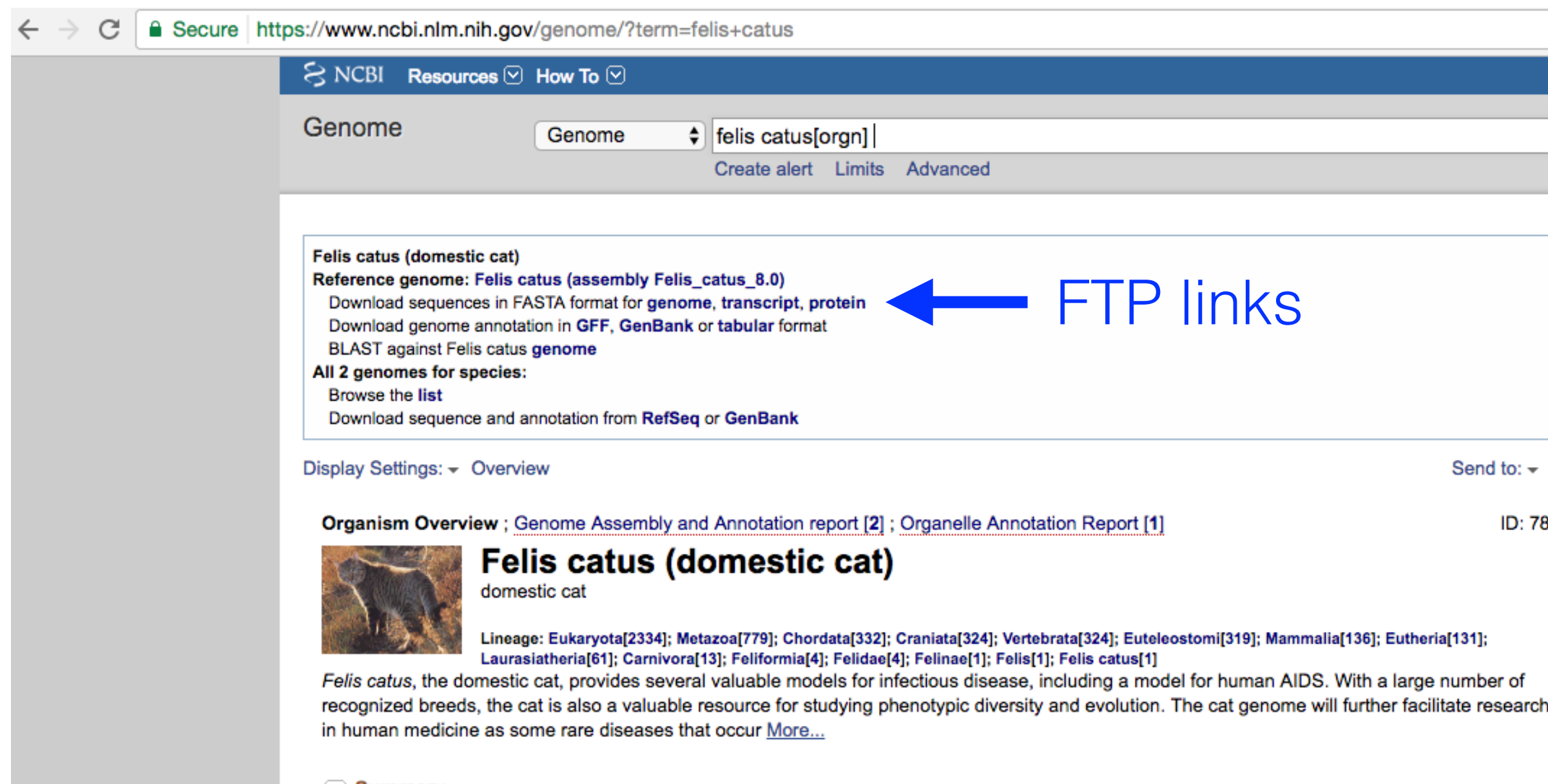
<ftp://ftp.ncbi.nlm.nih.gov/genomes/>

Index of /genomes/

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

You can download data from the command line

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




NCBI Resources How To

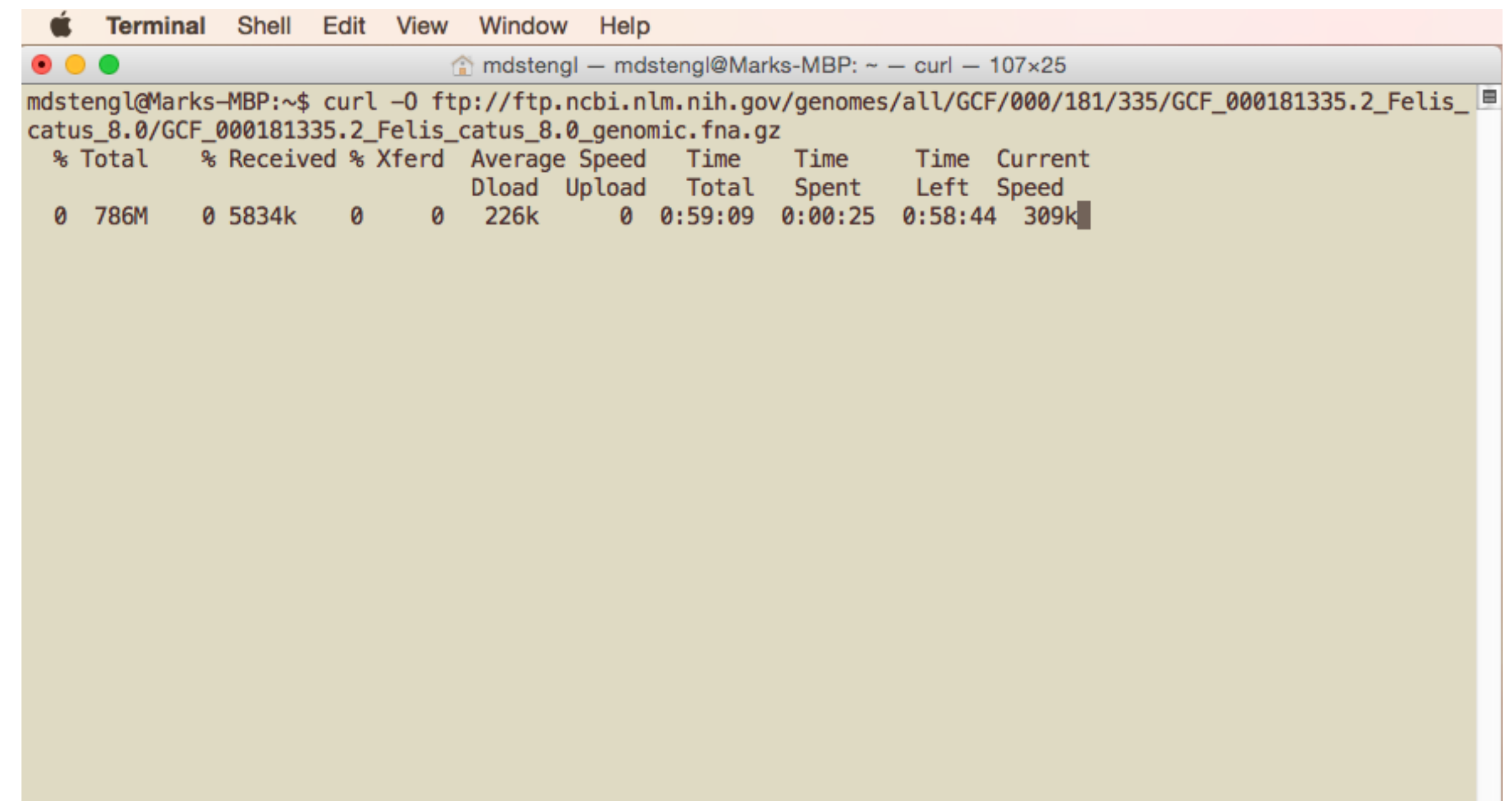
Genome Genome felis catus[orgn] Create alert Limits Advanced

Felis catus (domestic cat)
Reference genome: **Felis catus (assembly Felis_catus_8.0)**
Download sequences in FASTA format for **genome, transcript, protein** ← **FTP links**
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...](#)



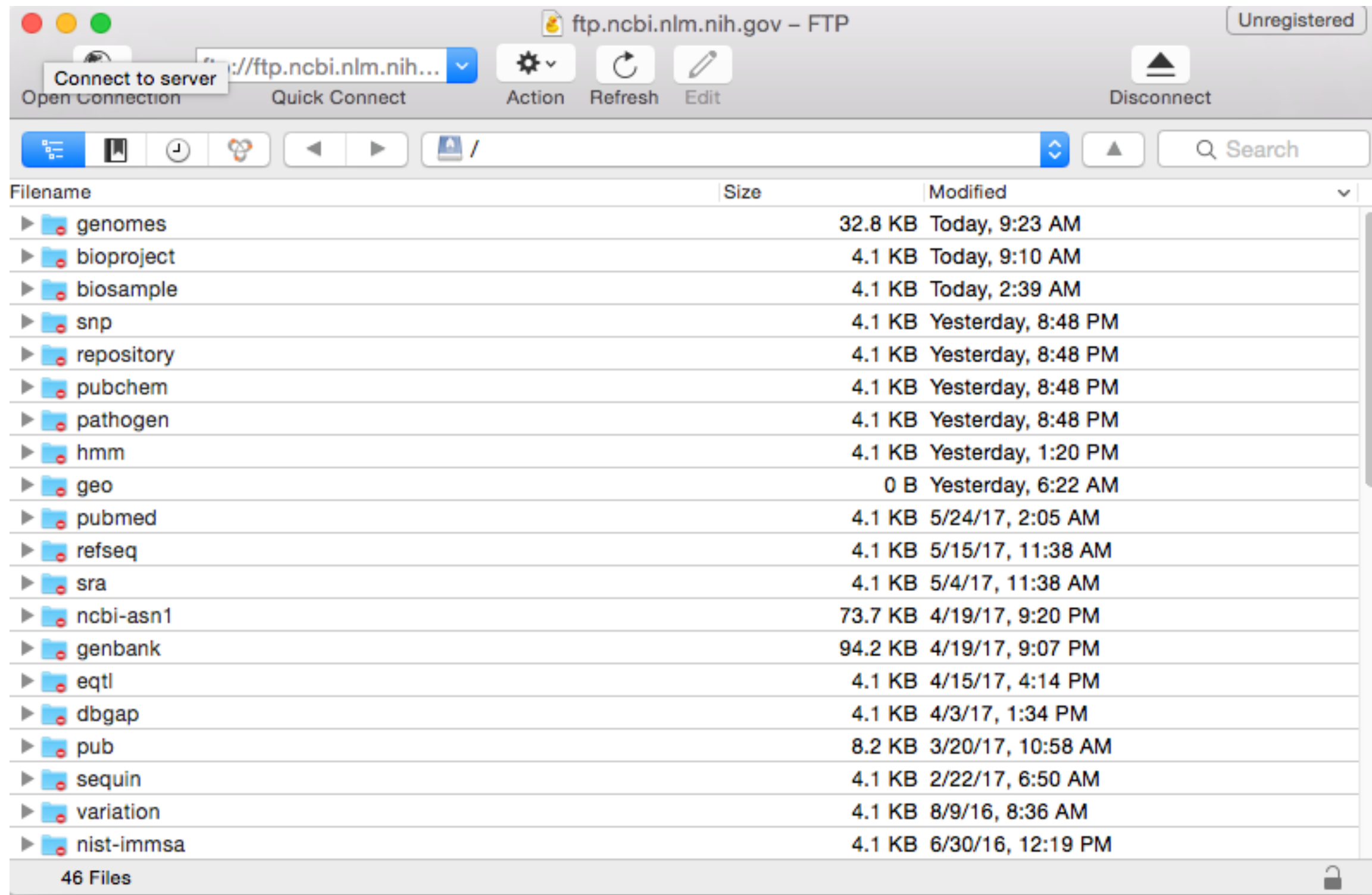
```
Terminal Shell Edit View Window Help
mdstengl — mdstengl@Marks-MBP: ~ — curl — 107x25
mdstengl@Marks-MBP:~$ curl -O ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/181/335/GCF_000181335.2_Felis_catus_8.0/GCF_000181335.2_Felis_catus_8.0_genomic.fna.gz
% Total    % Received % Xferd  Average Speed   Time    Time     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



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



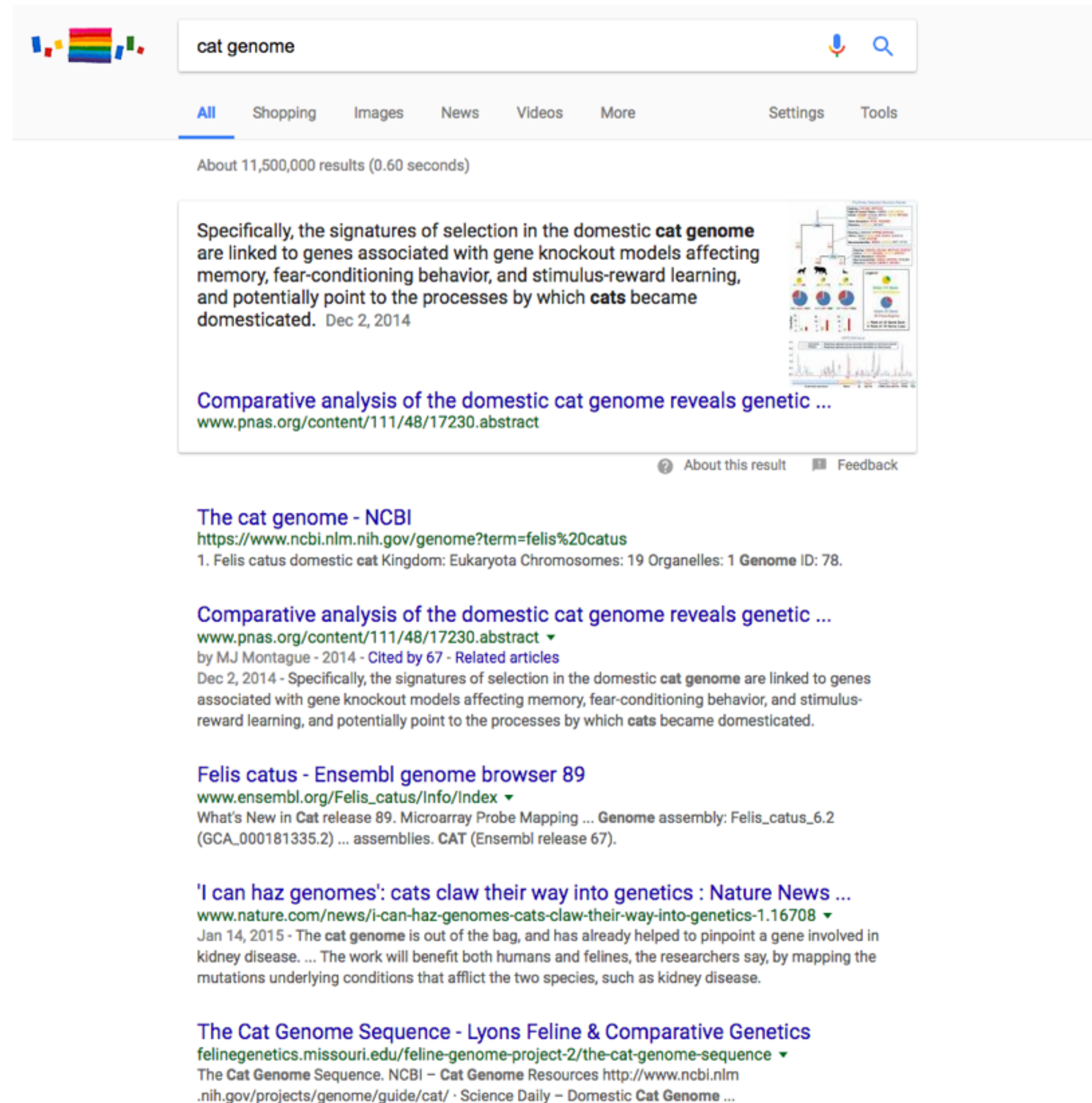
Genome browsers, like Ensembl and UCSC, offer additional functionality

The screenshot displays the Ensembl genome browser interface for the Cat (Felis catus 6.2) genome. The main view shows Chromosome A1: 11,397,075-11,456,215. The gene BRCA2 is highlighted in a red box, with its transcript BRCA2-201 also indicated. The interface includes a navigation menu on the left with options like 'Region in detail', 'Comparative Genomics', and 'Genetic Variation'. The main display area shows a genomic track with various annotations, including contigs, genes, and protein coding regions. A detailed view of the BRCA2 gene is shown below, with tracks for 40 way GERP elements, vertebrate UniProt proteins, cat cDNAs, vertebrate cDNAs, EST clusters, genes, contigs, EST clusters, vertebrate cDNAs, sequence variants (dbSNP and all other sources), and %GC. The BRCA2-201 transcript is shown in red, and the protein coding region is highlighted in green. The detailed view also shows the BRCA2-201 protein coding region in red and the BRCA2-201 transcript in green. The BRCA2-201 transcript is shown in red, and the protein coding region is highlighted in green. The BRCA2-201 transcript is shown in red, and the protein coding region is highlighted in green.

Genome browsers, like Ensembl and UCSC, offer additional functionality

The image shows a screenshot of the UCSC Genome Browser interface. At the top, there is a navigation bar with links for Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, View, Help, and About Us. The main heading reads "UCSC Genome Browser on Cat Nov. 2014 (ICGSC Felis_catus_8.0/felCat8) Assembly". Below this, there are navigation controls including "move" buttons (left, right, zoom in, zoom out) and a search box containing "chrA1:11,533,798-11,593,319 59,522 bp.". A scale bar at the top of the tracks shows a 20 kb region from 11,540,000 to 11,590,000 bp. The tracks include: Scale chrA1, Gap, Other RefSeq (with BRCR2 highlighted), RefSeq Genes, Cat mRNAs from GenBank (with AB107955 highlighted), Spliced ESTs, and Repeating Elements by RepeatMasker. A detailed instruction box at the bottom explains how to interact with the tracks, such as clicking for details, zooming in, and reordering tracks. A toolbar at the bottom right contains buttons for track search, default tracks, default order, hide all, add custom tracks, track hubs, configure, multi-region, reverse, resize, and refresh.

Finally, there's absolutely nothing wrong with using Google



A screenshot of a Google search for "cat genome". The search bar contains the text "cat genome" and has a microphone icon and a search icon to its right. Below the search bar are navigation tabs for "All", "Shopping", "Images", "News", "Videos", "More", "Settings", and "Tools". The "All" tab is selected. Below the tabs, it says "About 11,500,000 results (0.60 seconds)".

The first search result is a featured snippet with the following text: "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". To the right of this text is a small image showing a phylogenetic tree and several charts. Below the snippet is a link: "Comparative analysis of the domestic cat genome reveals genetic ..." with the URL "www.pnas.org/content/111/48/17230.abstract". Below the link are two small icons: "About this result" and "Feedback".

The second search result is "The cat genome - NCBI" with the URL "https://www.ncbi.nlm.nih.gov/genome?term=felis%20catus". Below the URL is the text "1. Felis catus domestic cat Kingdom: Eukaryota Chromosomes: 19 Organelles: 1 Genome ID: 78.".

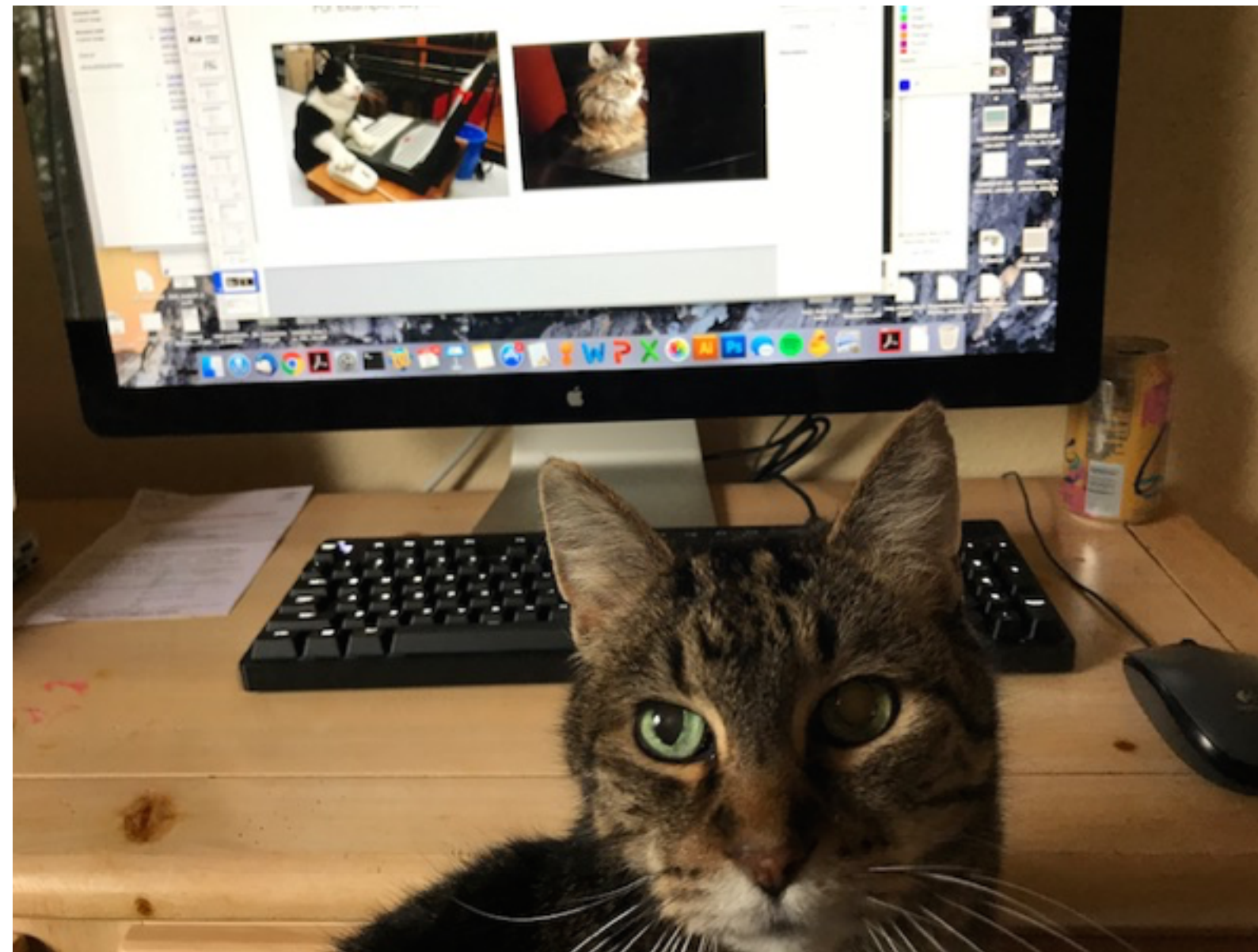
The third search result is "Comparative analysis of the domestic cat genome reveals genetic ..." with the URL "www.pnas.org/content/111/48/17230.abstract". Below the URL is the text "by MJ Montague - 2014 - Cited by 67 - Related articles" and "Dec 2, 2014 - Specifically, the signatures of selection in 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.".

The fourth search result is "Felis catus - Ensembl genome browser 89" with the URL "www.ensembl.org/Felis_catus/Info/Index". Below the URL is the text "What's New in Cat release 89. Microarray Probe Mapping ... Genome assembly: Felis_catus_6.2 (GCA_000181335.2) ... assemblies. CAT (Ensembl release 67).".

The fifth search result is "'I can haz genomes': cats claw their way into genetics : Nature News ..." with the URL "www.nature.com/news/i-can-haz-genomes-cats-claw-their-way-into-genetics-1.16708". Below the URL is the text "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 sixth search result is "The Cat Genome Sequence - Lyons Feline & Comparative Genetics" with the URL "felinegenetics.missouri.edu/feline-genome-project-2/the-cat-genome-sequence". Below the URL is the text "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