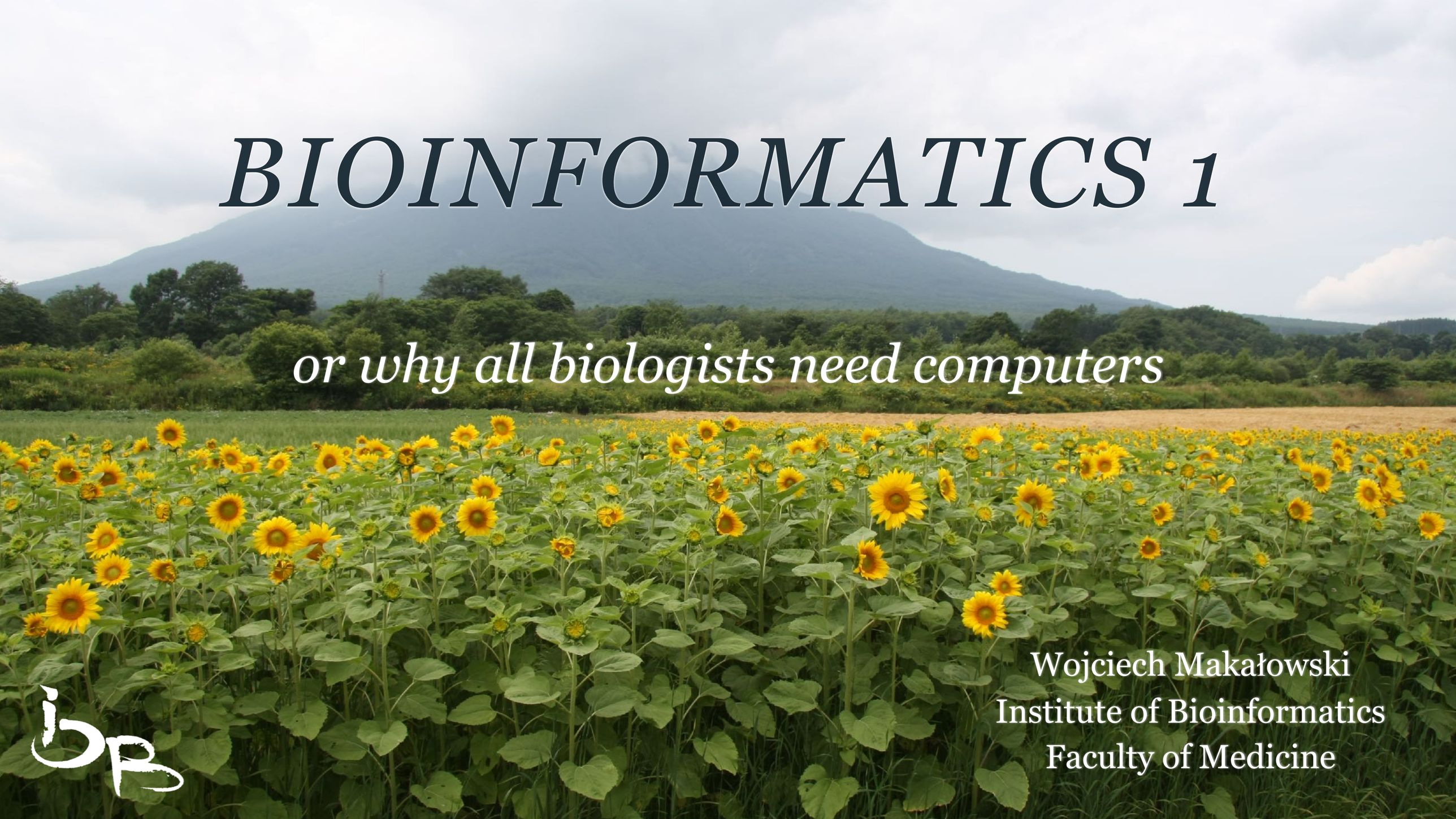
BIOINFORMATICS 1

or why all biologists need computers



Wojciech Makałowski
Institute of Bioinformatics
Faculty of Medicine

TOPICS TO BE COVERED IN THIS COURSE

- Introduction to bioinformatics from the evolutionary perspective. [WM]
- Introduction to Sequence Analysis. [WM]
- Genome Annotation. [WM]
- Phylogenetic inference. [WM]
- Differential gene expression. [JG]
- Introduction to system biology. [EK]
- Introduction to artificial inteligence for biologists [XJ]

HANDS ON COMPUTER LAB

This year via zoom only

- First session
  - from BLAST to phylogenetic inference (week of November 27, 2023)
- Second session
  - transcriptome analyses (week of December 4, 2023)
- Registration to practicals will be open in early November

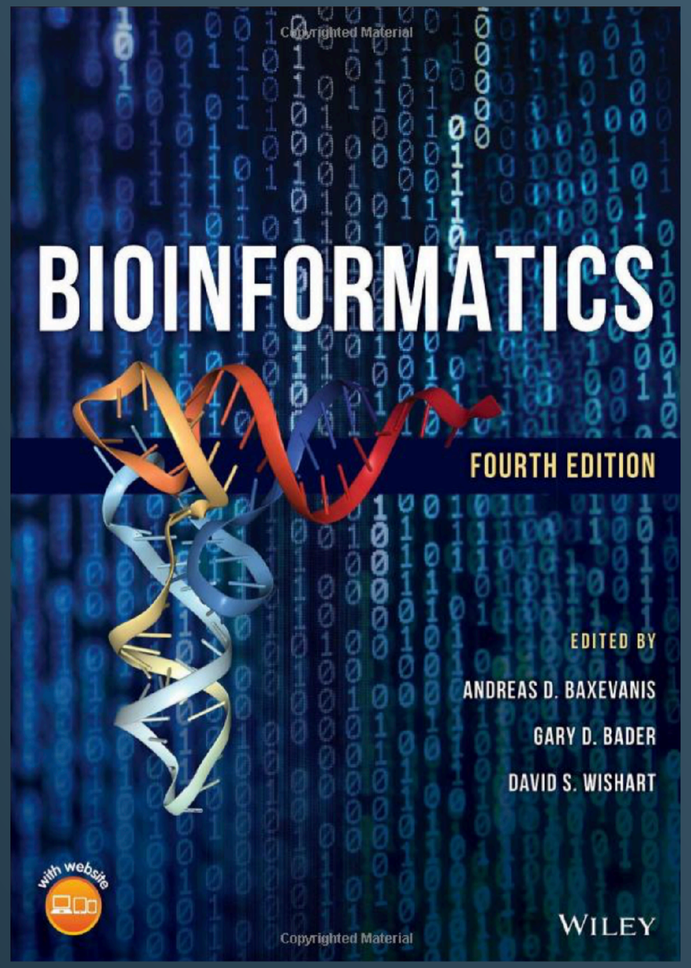
CONTACT

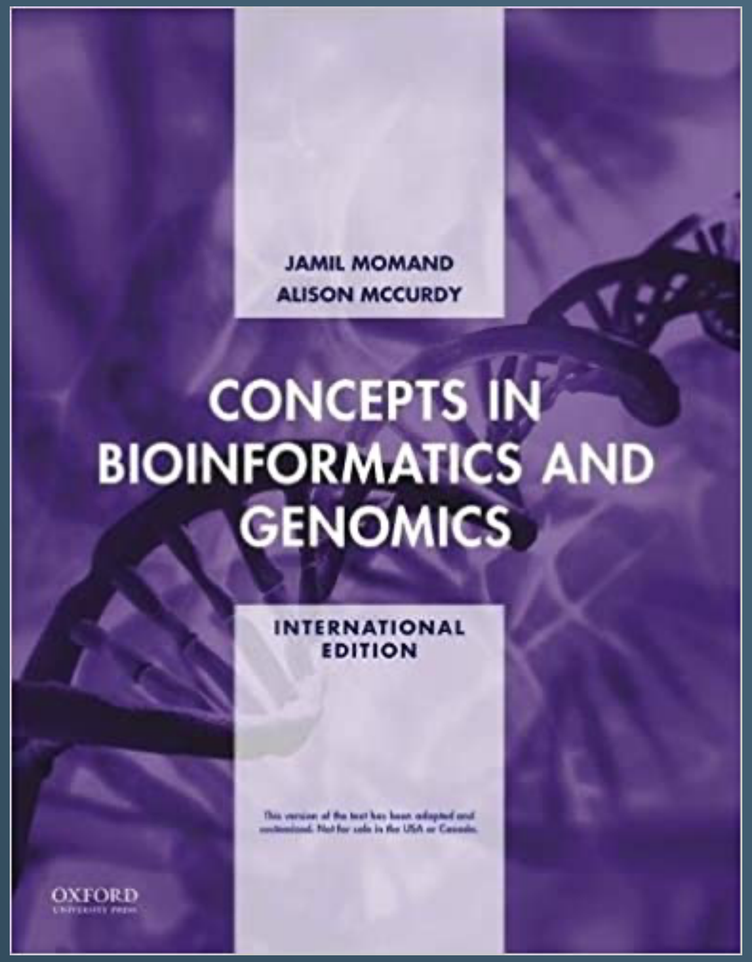
- Prof. Dr. Wojciech Makałowski wojmak@uni-muenster.de
- Prof. Dr. Jürgen R. Gadau gadauj@uni-muenster.de
- Prof. Dr. Eberhard Korsching eberhard.Korsching@uni-muenster.de
- Prof. Dr. Xiaoyi Jiang xjiang@uni-muenster.de
- Mr. Felix Menske f_mans02@uni-muenster.de (lab coordinator)
- http://bioinformatics.uni-muenster.de/teaching/Current/bioinf1/index.hbi
- office hours - by appointment

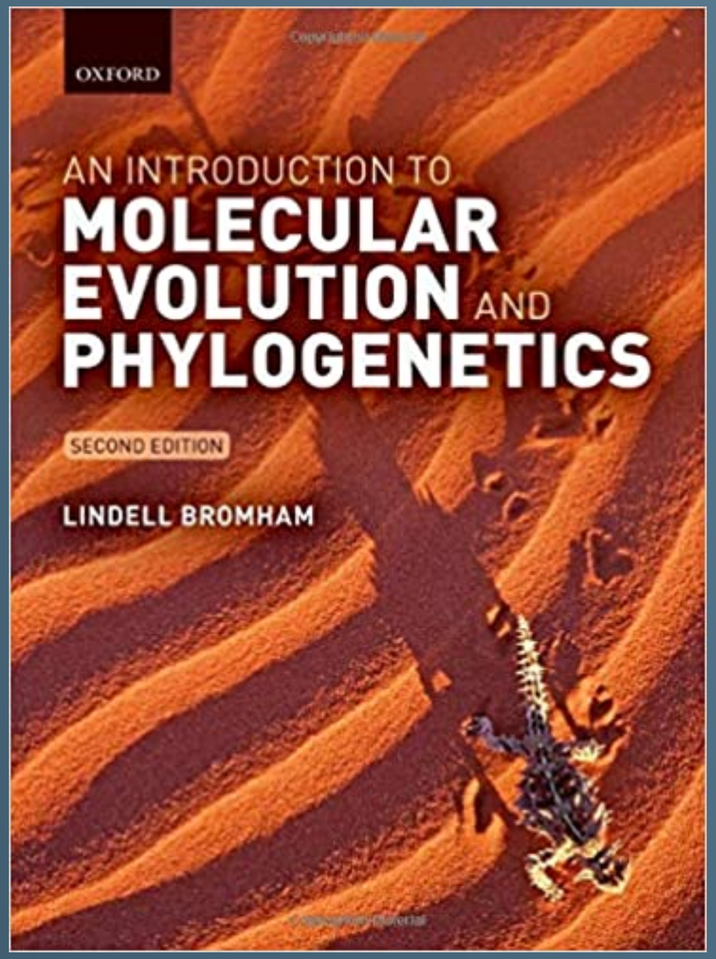
GRADING

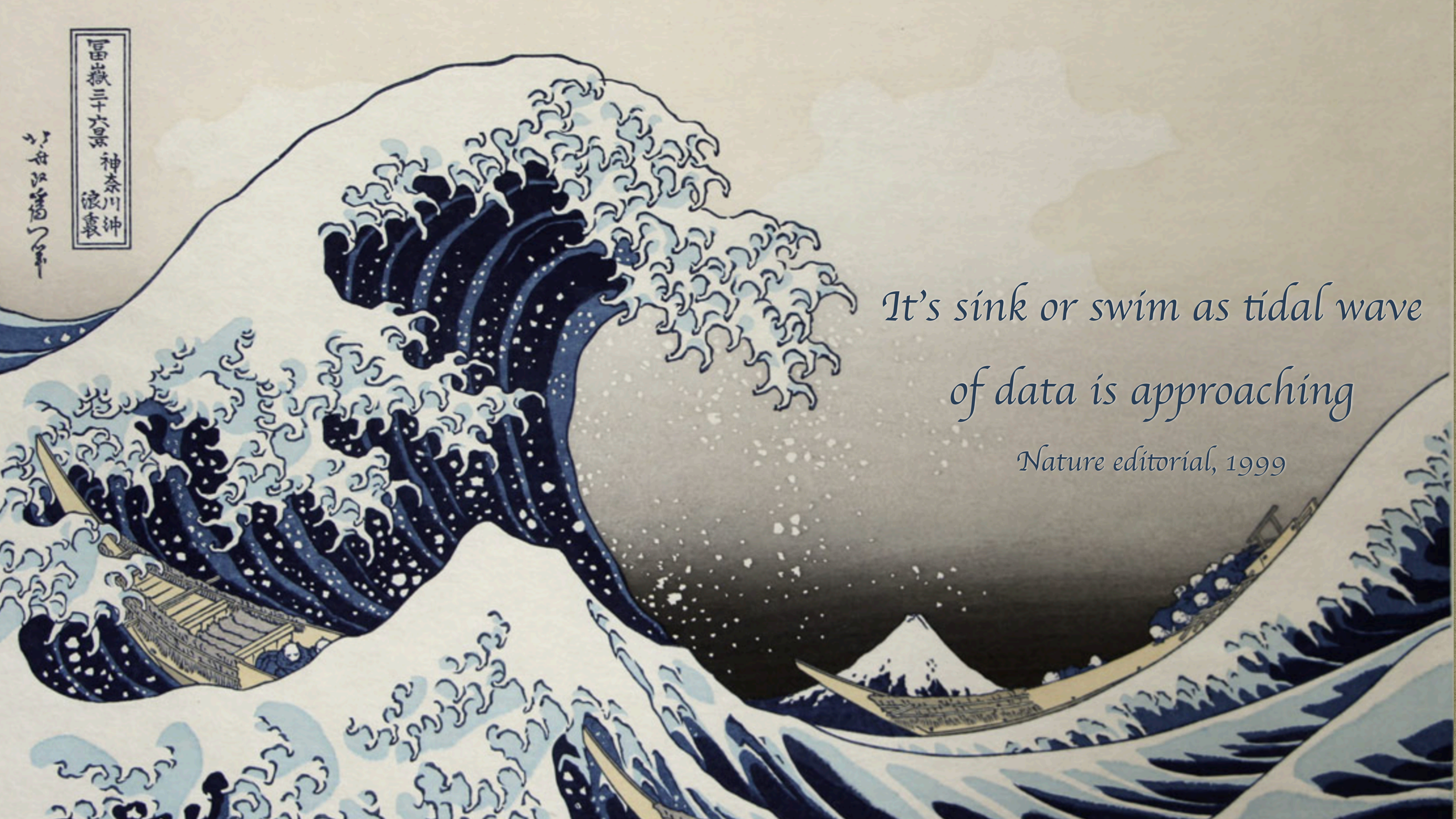
- 60% of the final grade comes from the exam
  - date and form to be determined
- 40% of the final grade based on practicals

RECOMMENDED BOOKS









It's sink or swim as tidal wave

of data is approaching

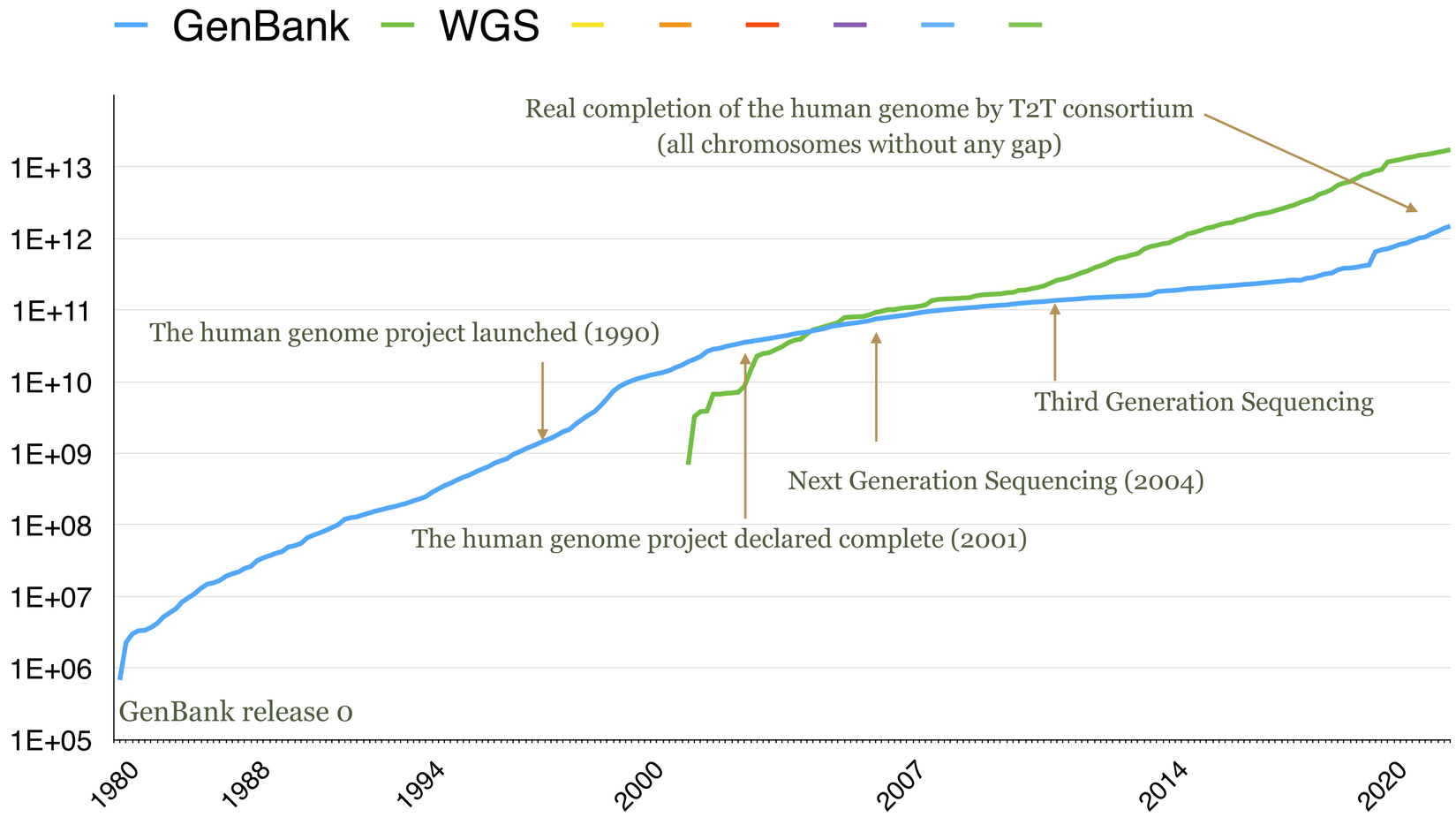
Nature editorial, 1999


Unfortunately, it is not a tidal wave,
it is a tsunami!

GROWTH OF BIOMEDICAL INFORMATION - GENBANK

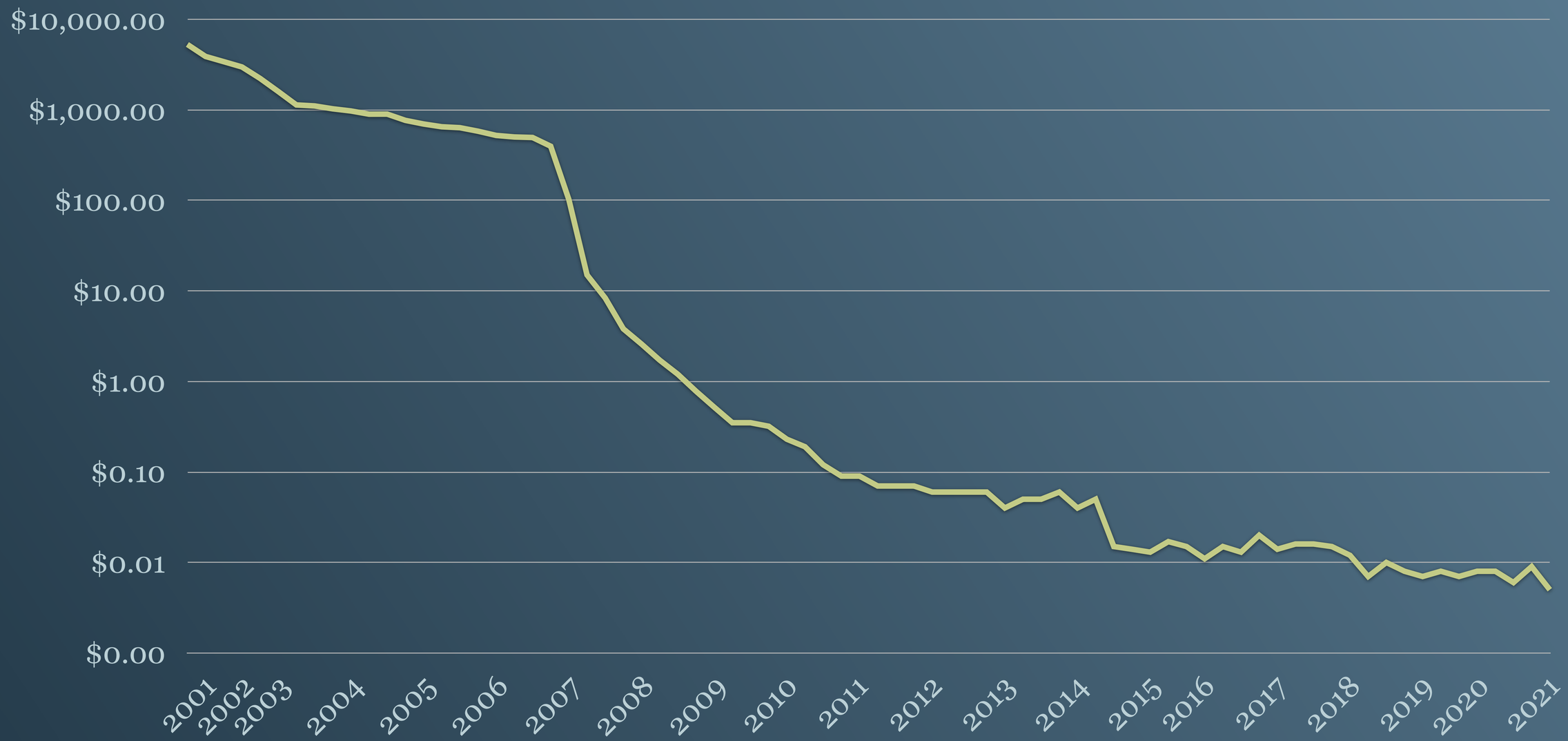
- GenBank Rel. 0 (May 1980)
  - 1000 seq
  - 100,000 nt

- GenBank Rel. 251 (August 2023)
  - 246 mln seq; 2.1 trillion nt
  - 22.3 trillion nt in the “whole genome shotgun” section

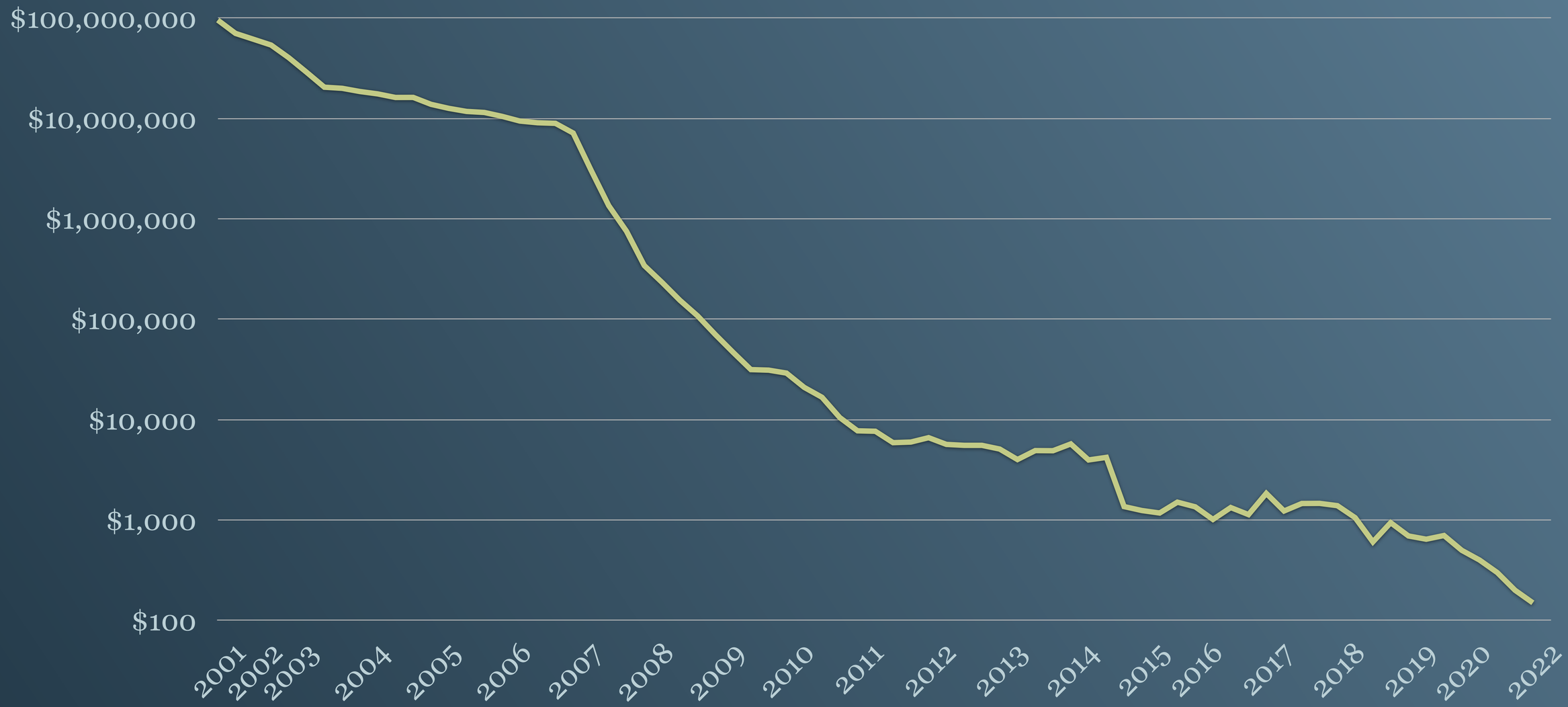


https://www.ncbi.nlm.nih.gov/genbank/statistics/

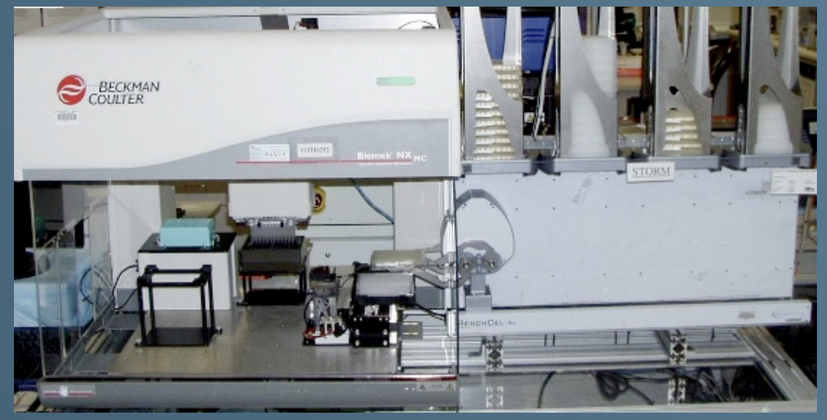
SEQUENCING COST PER MEGABASE



SEQUENCING COST PER HUMAN GENOME



TECHNOLOGY MEETS BIOLOGY



GETTING SEQUENCES

TGCATCGATCGTAGCTAGCTAGCGCATGCTAGCTAGCTAGCTAGCTACGATGCATCG
TGCATCGATCGATGCATGCTAGCTAGCTAGCTAGCATGCTAGCTAGCTAGCTATTGG
CGCTAGCTAGCATGCATGCATGCATCGATGCATCGATTATAAGCGCGATGACGTCAG
CGCGCGCATTATGCCGCGGCATGCTGCGCACACACAGTACTATAGCATTAGTAAAAA
GGCCGCGTATATTTTACACGATAGTGCGGCGCGGCGCGTAGCTAGTGCTAGCTAGTC
TCCGGTTACACAGGTAGCTAGCTAGCTGCTAGCTAGCTGCTGCATGCATGCATTAGT
AGCTAGTGTAGCTAGCTAGCATGCTGCTAGCATGCAGCATGCATCGGGCGCGATGCT
GCTAGCGCTGCTAGCTAGCTAGCTAGCTAGGCGCTAATTATTTATTTTGGGGGGTTA
AAAAAAAAAATTTCGCTGCTTATACCCCCCCCCACATGATGATCGTTAGTAGCTACT
AGCTCTCATCGCGCGGGGGGATGCTTAGCGTGGTGTGTGTGTGTGGTGTGTGTGGTC
CTATAATTAGTGCATCGGCGCATCGATGGCTAGTCGATCGATCGATTTTATATATCT
AAAGACCCCATCTCTCTCTCTTTTCCCTTCTCTCGCTAGCGGGCGGTACGATTTACC
GGCCGCGTATATTTTACACGATAGTGCGGCGCGGCGCGTAGCTAGTGCTAGCTAGTC
AGCTCTCATCGCGCGGGGGGATGCTTAGCGTGGTGTGTGTGTGTGGTGTGTGTGGTC
TGCATCGATCGATGCATGCTAGCTAGCTAGCTAGCATGCTAGCTAGCTAGCTATTGG
CTATAATTAGTGCATCGGCGCATCGATGGCTAGTCGATCGATCGATTTTATATATCT
CGCTAGCTAGCATGCATGCATGCATCGATGCATCGATTATAAGCGCGATGACGTCAG
TCCGGTTACACAGGTAGCTAGCTAGCTGCTAGCTAGCTGCTGCATGCATGCATTAGT

READING ≠ UNDERSTANDING

Carmina qui quondam studio florente peregi, flebilis heu maestos cogor inire modos.

Ecce mihi lacerae dictant scribenda Camenae et ueris elegi fletibus ora rigant.

Boethius, *Consolatio Philosophiae*

READING ≠ UNDERSTANDING

We shall best understand the probable course of natural selection by taking the case of a country undergoing some physical change. If the country were open were open on its borders, new forms would certainly immigrate, and this also would bla, bla bla become extinct inhabitants.

Charles Darwin - *The Origin of Species*

CHALLENGE: HOW FROM THIS…

TGCATCGATCGTAGCTAGCTAGCGCATGCTAGCTAGCTAGCTAGCTACGATGCATCG
TGCATCGATCGATGCATGCTAGCTAGCTAGCTAGCATGCTAGCTAGCTAGCTATTGG
CGCTAGCTAGCATGCATGCATGCATCGATGCATCGATTATAAGCGCGATGACGTCAG
CGCGCGCATTATGCCGCGGCATGCTGCGCACACACAGTACTATAGCATTAGTAAAAA
GGCCGCGTATATTTTACACGATAGTGCGGCGCGGCGCGTAGCTAGTGCTAGCTAGTC
TCCGGTTACACAGGTAGCTAGCTAGCTGCTAGCTAGCTGCTGCATGCATGCATTAGT
AGCTAGTGTAGCTAGCTAGCATGCTGCTAGCATGCAGCATGCATCGGGCGCGATGCT
GCTAGCGCTGCTAGCTAGCTAGCTAGCTAGGCGCTAATTATTTATTTTGGGGGGTTA
AAAAAAAAAATTTCGCTGCTTATACCCCCCCCCCACATGATGATCGTTAGTAGCTACT
AGCTCTCATCGCGCGGGGGGGATGCTTAGCGTGGTGTGTGTGTGTGGTGTGTGTGGTC
CTATAATTAGTGCATCGGCGCATCGATGGCTAGTCGATCGATCGATTTTATATATCT
AAAGACCCCATCTCTCTCTCTTTTCCCTTCTCTCGCTAGCGGGCGGTACGATTTACC
GGCCGCGTATATTTTACACGATAGTGCGGCGCGGCGCGTAGCTAGTGCTAGCTAGTC
AGCTCTCATCGCGCGGGGGGGATGCTTAGCGTGGTGTGTGTGTGTGGTGTGTGTGGTC
TGCATCGATCGATGCATGCTAGCTAGCTAGCTAGCATGCTAGCTAGCTAGCTATTGG
CTATAATTAGTGCATCGGCGCATCGATGGCTAGTCGATCGATCGATTTTATATATCT
CGCTAGCTAGCATGCATGCATGCATCGATGCATCGATTATAAGCGCGATGACGTCAG
TCCGGTTACACAGGTAGCTAGCTAGCTGCTAGCTAGCTGCTGCATGCATGCATTAGT


infer this

HOW TO SOLVE A PROBLEM - A HUMAN OR A COMPUTER?

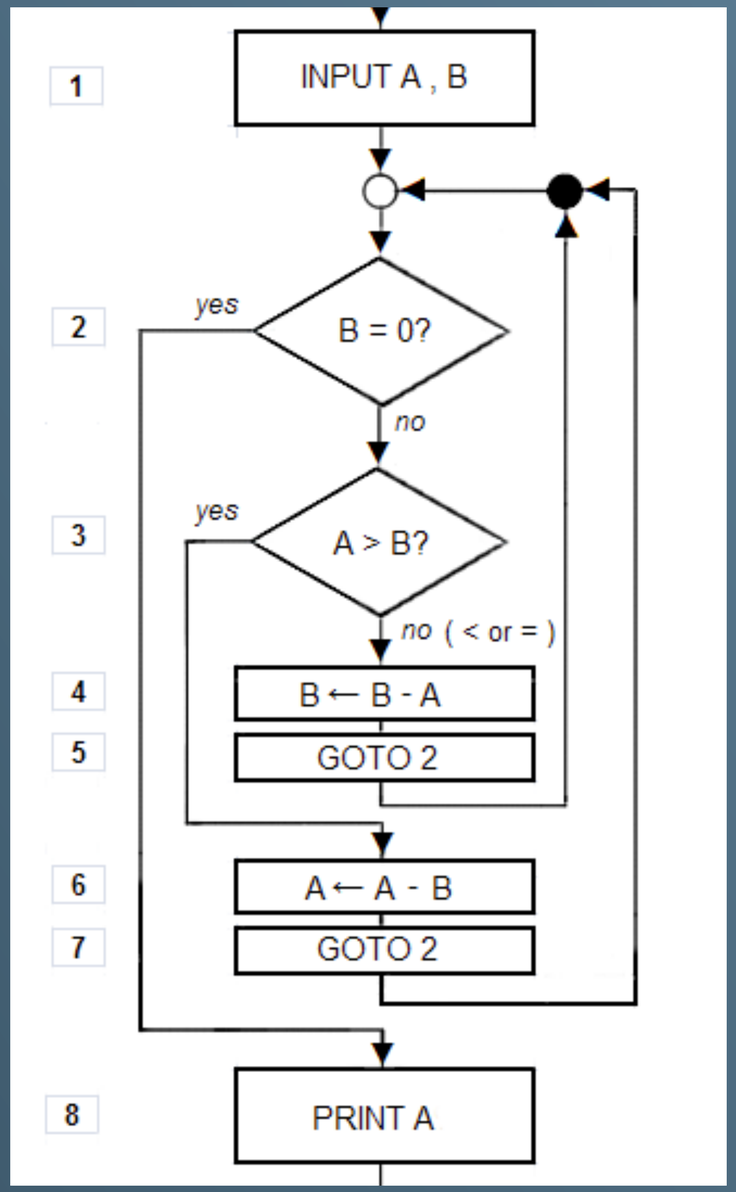
- very smart
- slow
- error prone
- doesn’t like repetitive tasks

- not so smart (stupid)
- extremely fast
- very accurate
- doesn’t understand human languages;

  needs instruction provided in a special way

ALGORITHM

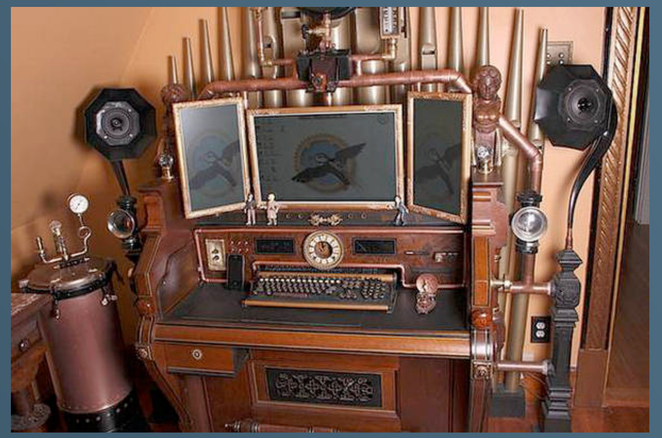
A step-by-step problem-solving procedure, especially an established, recursive computational procedure for solving a problem in a finite number of steps.



EXAMPLE TASK: PUT SHOES ON!

A human just understands an order and often executes it automatically even without thinking

A computer needs detailed instruction (an algorithm)



PUT SHOES ON!
INSTRUCTION FOR A COMPUTER

1. Find two the same shoes
2. Check if you have left and right shoe
3. Check if they are of the same size
4. Check if this is the right size
5. Put the left shoe on
6. Put the right shoe on
7. Tie the laces

THE ORIGIN OF THE FIELD

Paulien Hogeweg coined the term *bioinformatica* to define “the study of informatic processes in biotic systems.”

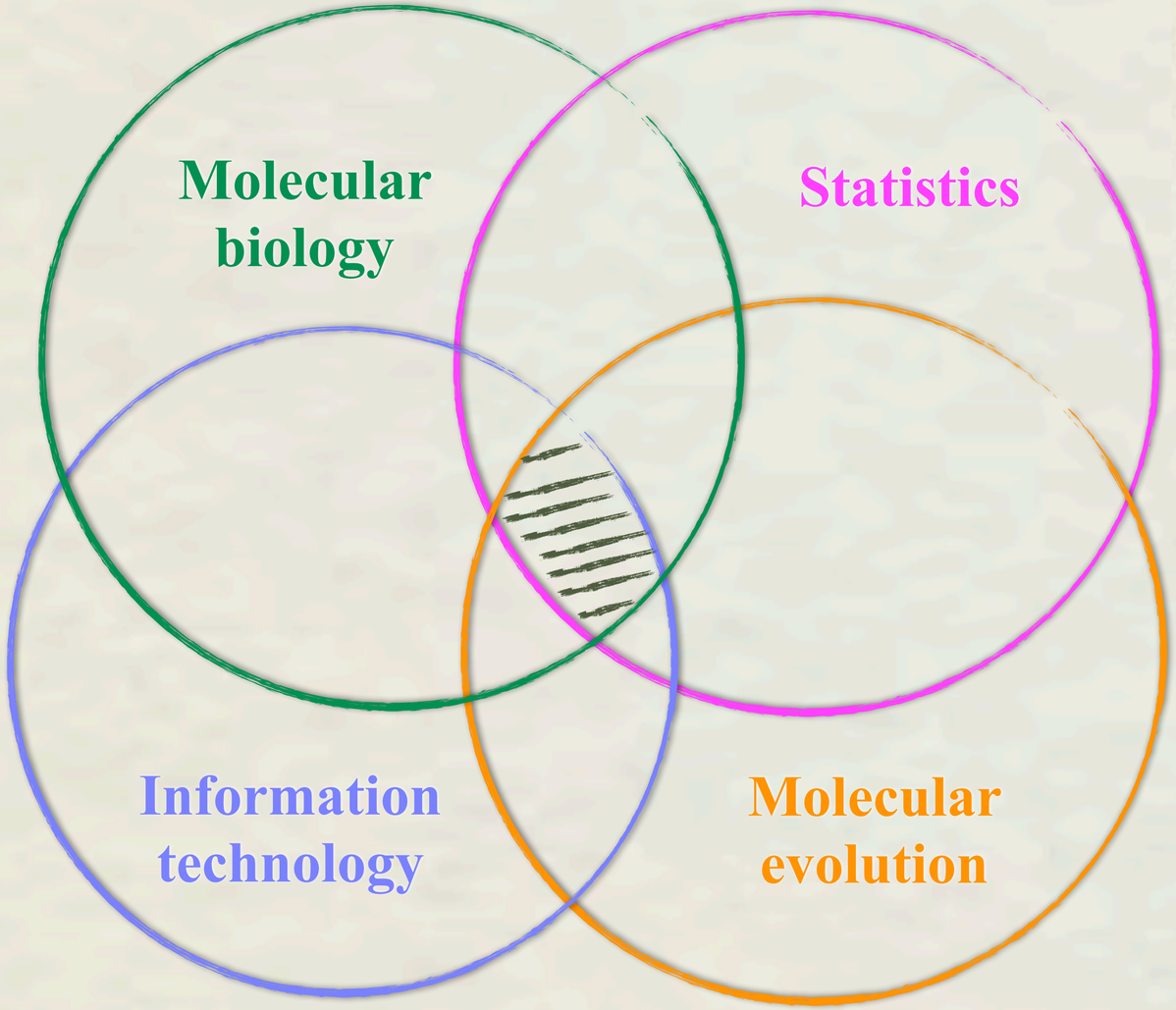
Hesper B, Hogeweg P (1970) Bioinformatica: een werkconcept. Kameleon 1(6): 28–29. (In Dutch.) Leiden: Leidse Biologen Club.

… but its origin can be tracked back many decades earlier.

BIOINFORMATICS EMERGED AS AN INTERSECTION BETWEEN DIFFERENT DISCIPLINES

James Watson

Alan Turing



Thomas Bayes

Motoo Kimura

BIOINFORMATICS - DEFINITION

- Research, development, or application of computational tools and approaches for expanding the use of biological data, including those to acquire, store, organize, archive, analyze, or visualize such data.

- Its goal is to enable biological discovery based on existing information or in other words transform biological data into information and eventually into knowledge.

ROLE OF BIOINFORMATICS IN MODERN LIFE SCIENCES

molecular biology

molecular evolution

genomics

system biology

protein engineering

drug design

human genetics

personalized medicine

biogeography

you name it…

TWO MAJOR CHALLENGES

How to store the data?

How to analyze the data?

HOW TO STORE DATA

Ad hoc storage

Organized storage

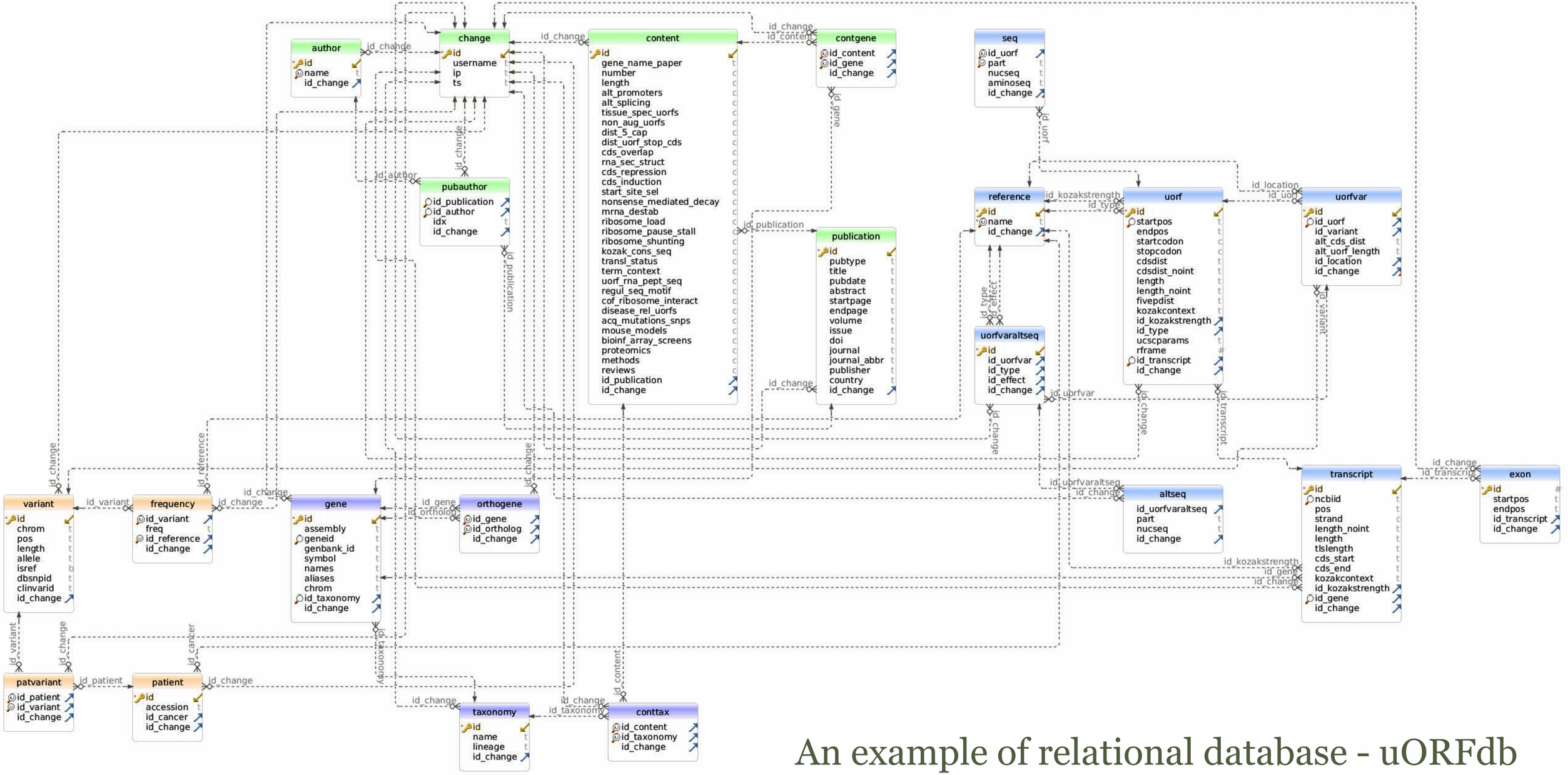
BIOLOGICAL DATABASES

BIOLOGICAL DATABASES

- organized sets of large amount of data, usually coupled with a software that enables data search, information extraction, and data update
- databases should be characterized by
  - easy data access
  - the possibility to extract only the information that is desirable

MODERN RESOURCES

- Relational Database Management Systems (RDBMS)
  - Introduced in the 1970s
  - Off-the-shelf software (commercial and open source)
    - Oracle, DB2, MySQL, PostgradeSQL
  - High level declarative language - SQL
  - Concurrency
  - Transaction control
  - Consistency



An example of relational database - uORFdb

https://bioinformatics.uni-muenster.de/tools/uorfdb/index.pl

PRIMARY VS. SECONDARY DATABASE

Primary db

- Strictly repository database
- Original submitter “controls” a record, including updating information
- Database administrator can validate and check data for consistency
- Example: GenBank

Secondary db

- Original data, for example sequences, are post-processed adding more biological information
- Usually more specialized than repository databases
- Curators of the database take control of the content
- Example: uORFdb

CRITICAL ISSUES FOR BIOLOGICAL DATABASES

- Annotation
  - Correctness
  - Consistency
  - Quality
- Archival Quality
- Updates
  - Raw data
  - Annotation

CRITICAL ISSUES - ANNOTATION

- Correctness – many genes are annotated primarily based on sequence comparisons. Annotation is copied from a similar sequence to a novel sequence. This may cause some problems
  - Comparison may have been done when the data was less complete
  - If a sequence is incorrectly annotated, this error propagates through the database

CRITICAL ISSUES ANNOTATION QUALITY

- Who supplies the annotation? An expert, or a non-expert at the database
- Many databases have defined groups of “experts” to help annotated genes or gene families, but there is no peer review of information in databases
- What is the vocabulary?

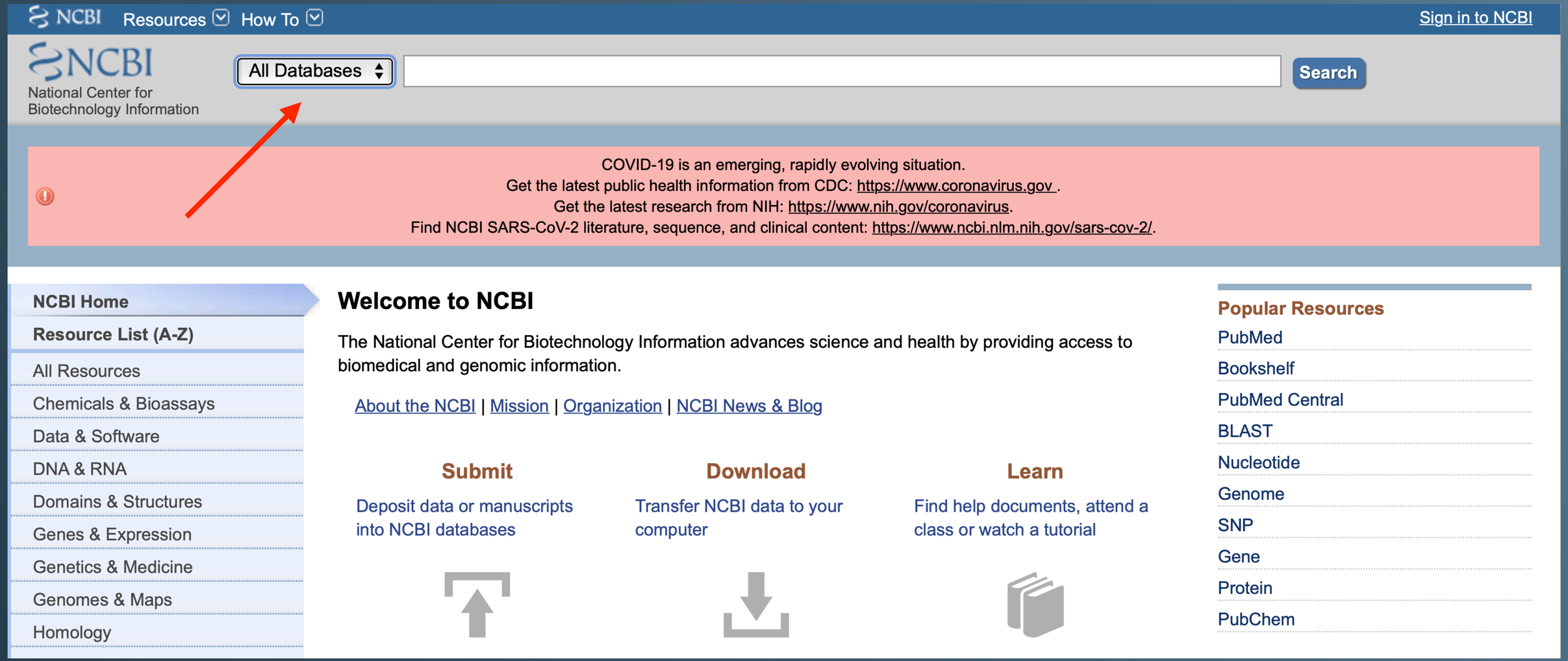
CRITICAL ISSUES ARCHIVAL QUALITY

- Databases have been torn between trying to be archival – to simply report information as experts publish it (*primary databeses*), or curated – to provide the best editorially reviewed data on a topic (*secondary DB*).
- Can the same entry be recovered later?
  - Accession numbers are more stable than entry or locus names
  - Many databases do not note that there have been changes to the data! What you retrieve today may be different than yesterday

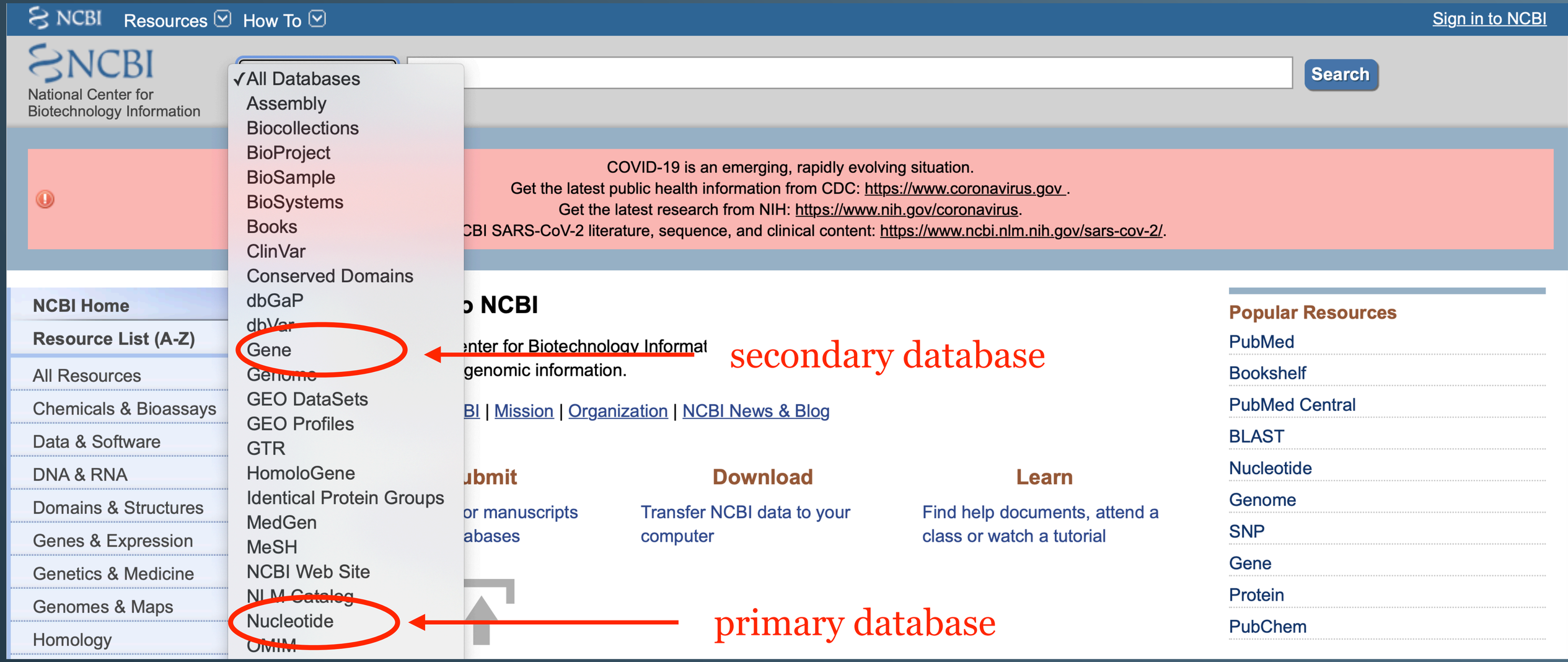
CRITICAL ISSUES: UPDATES

- How often are updates done? Major databases take direct submissions.
- Generally, only the original submitter can change an entry, even if you can prove it is wrong. This is tied to the question of archival versus curated.
- How is annotation updated as more knowledge is available? Who decides?

NCBI - HOME OF MANY IMPORTANT DATABASES



NCBI - HOME OF MANY IMPORTANT DATABASES



GenBank - A PRIMARY NUCLEOTIDE SEQUENCE DATABASE

- GenBank ® is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences

- GenBank is part of the International Nucleotide Sequence Database Collaboration, which comprises the DNA DataBank of Japan (DDBJ), the European Nucleotide Archive (ENA), and GenBank at NCBI

- These three organizations exchange data on a daily basis

```
LOCUS       MF084995                570 bp    mRNA    linear   INV 08-JAN-2018
DEFINITION  Acanthamoeba castellanii Pam18 isoform A mRNA, complete cds.
ACCESSION   MF084995
VERSION     MF084995.1
KEYWORDS    .
SOURCE      Acanthamoeba castellanii
  ORGANISM  Acanthamoeba castellanii
            Eukaryota; Amoebozoa; Discosea; Longamoebia; Centramoebida;
            Acanthamoebidae; Acanthamoeba.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Wojtkowska,M., Buczek,D., Suzuki,Y., Shabardina,V., Makalowski,W.
            and Kmita,H.
  TITLE     The emerging picture of the mitochondrial protein import complexes
            of Amoebozoa supergroup
  JOURNAL   BMC Genomics 18 (1), 997 (2017)
   PUBMED   29284403
  REMARK    Publication Status: Online-Only
REFERENCE   2  (bases 1 to 570)
  AUTHORS   Buczek,D., Wojtkowska,M., Suzuki,Y., Makalowski,W. and Kmita,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2017) Adam Mickiewicz University, Institute of
            Molecular Biology and Biotechnology, Umultowska 89, Poznan 61-614,
            Poland
COMMENT     ##Assembly-Data-START##
            Assembly Method       :: Trinity De Novo Assembly v. 2012.10.05
            Sequencing Technology :: Illumina
            ##Assembly-Data-END##
FEATURES             Location/Qualifiers
     source          1..570
                     /organism="Acanthamoeba castellanii"
                     /mol_type="mRNA"
                     /db_xref="taxon:5755"
     CDS             113..406
                     /note="Pam18-A"
                     /codon_start=1
                     /product="Pam18 isoform A"
                     /protein_id="AUI80419.1"
                     /translation="MNAYKHFKAGNLTLPKGMVPKGPSRMKSYYTGGFESEMTRAEAA
                     LILSVRQGASKEKIKMAHRRIMLANHPDNGGSDYVASKVNEAKDLLLKDLGDD"
ORIGIN
        1 ccggaacttg catctgcgac catctgccgc tgtatcgatc cgggagcaat tggctactct
       61 tcttgtggcg ggtgtggtga ttgctggtgt ggccatcggt gggcgcgtgg ctatgaacgc
      121 ctacaagcac ttcaaggccg gcaacttgac cctccccaag ggcatggtgc ccaagggtcc
      181 atcgaggatg aagtcctact acacgggcgg ctttgagtcg gagatgaccc gcgccgaagc
      241 cgctctcatc ctcagtgtcc gacaaggcgc ctcgaaggag aagatcaaga tggcccacag
      301 gcggatcatg ttggcgaacc atcccgacaa tggaggcagc gactacgtgg cgtcgaaggt
      361 gaacgaagcc aaggacctgc tgctcaagga tctcggcgac gactgaggcc cctcttccaa
      421 aataacgaga gcaccaaaag gaccaccccc accaaacaca gacacacacc gagacaccac
      481 acaccgccgc cgccgccgcc accagagaga tcggggctgt acagagtact acaccaccat
      541 atatgcacac ccatattacc aacaaaaaaa
//
```

GenBank record example

Hyperlinked features can bring you to another NCBI database such as

- taxonomy
- PubMed
- protein

or

extract a particular feature from the record

```
LOCUS       MF084995                570 bp    mRNA    linear   INV 08-JAN-2018
DEFINITION  Acanthamoeba castellanii Pam18 isoform A mRNA, complete cds.
ACCESSION   MF084995
VERSION     MF084995.1
KEYWORDS    .
SOURCE      Acanthamoeba castellanii
  ORGANISM  Acanthamoeba castellanii
            Eukaryota; Amoebozoa; Discosea; Longamoebia; Centramoebida;
            Acanthamoebidae; Acanthamoeba.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Wojtkowska,M., Buczek,D., Suzuki,Y., Shabardina,V., Makalowski,W.
            and Kmita,H.
  TITLE     The emerging picture of the mitochondrial protein import complexes
            of Amoebozoa supergroup
  JOURNAL   BMC Genomics 18 (1), 997 (2017)
   PUBMED   29284403
  REMARK    Publication Status: Online-Only
REFERENCE   2  (bases 1 to 570)
  AUTHORS   Buczek,D., Wojtkowska,M., Suzuki,Y., Makalowski,W. and Kmita,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2017) Adam Mickiewicz University, Institute of
            Molecular Biology and Biotechnology, Umultowska 89, Poznan 61-614,
            Poland
COMMENT     ##Assembly-Data-START##
            Assembly Method       :: Trinity De Novo Assembly v. 2012.10.05
            Sequencing Technology :: Illumina
            ##Assembly-Data-END##
FEATURES             Location/Qualifiers
     source          1..570
                     /organism="Acanthamoeba castellanii"
                     /mol_type="mRNA"
                     /db_xref="taxon:5755"
     CDS             113..406
                     /note="Pam18-A"
                     /codon_start=1
                     /product="Pam18 isoform A"
                     /protein_id="AUI80419.1"
                     /translation="MNAYKHFKAGNLTLPKGMVPKGPSRMKSYYTGGFESEMTRAEAA
                     LILSVRQGASKEKIKMAHRRIMLANHPDNGGSDYVASKVNEAKDLLLKDLGDD"
ORIGIN
        1 ccggaacttg catctgcgac catctgccgc tgtatcgatc cgggagcaat tggctactct
       61 tcttgtggcg ggtgtggtga ttgctggtgt ggccatcggt gggcgcgtgg ctatgaacgc
      121 ctacaagcac ttcaaggccg gcaacttgac cctccccaag ggcatggtgc ccaagggtcc
      181 atcgaggatg aagtcctact acacgggcgg ctttgagtcg gagatgaccc gcgccgaagc
      241 cgctctcatc ctcagtgtcc gacaaggcgc ctcgaaggag aagatcaaga tggcccacag
      301 gcggatcatg ttggcgaacc atcccgacaa tggaggcagc gactacgtgg cgtcgaaggt
      361 gaacgaagcc aaggacctgc tgctcaagga tctcggcgac gactgaggcc cctcttccaa
      421 aataacgaga gcaccaaaag gaccaccccc accaaacaca gacacacacc gagacaccac
      481 acaccgccgc cgccgccgcc accagagaga tcggggctgt acagagtact acaccaccat
      541 atatgcacac ccatattacc aacaaaaaaa
//
```

BMC Genomics. 2017 Dec 29;18(1):997. doi: 10.1186/s12864-017-4383-1.

The emerging picture of the mitochondrial protein import complexes of Amoebozoa supergroup

Małgorzata Wojtkowska [1], Dorota Buczek [2 3], Yutaka Suzuki [4], Victoria Shabardina [3], Wojciech Makałowski [3], Hanna Kmita [2]

Affiliations + expand

PMID: 29284403 PMCID: PMC5747110 DOI: 10.1186/s12864-017-4383-1

Free PMC article

Abstract

Background: The existence of mitochondria-related organelles (MROs) is proposed for eukaryotic organisms. The Amoebozoa includes some organisms that are known to have mitosomes but also organisms that have aerobic mitochondria. However, the mitochondrial protein apparatus of this supergroup remains largely unsampled, except for the mitochondrial outer membrane import complexes studied recently. Therefore, in this study we investigated the mitochondrial inner membrane and intermembrane space complexes, using the available genome and transcriptome sequences.

Results: When compared with the canonical cognate complexes described for the yeast Saccharomyces cerevisiae, amoebozoans with aerobic mitochondria, display lower differences in the number of subunits predicted for these complexes than the mitochondrial outer membrane complexes, although the predicted subunits appear to display different levels of diversity in regard to phylogenetic position and isoform numbers. For the putative mitosome-bearing amoebozoans, the number of predicted subunits suggests the complex elimination distinctly more pronounced than in the case of the outer membrane ones.

Conclusion: The results concern the problem of mitochondrial and mitosome protein import machinery structural variability and the reduction of their complexity within the currently defined supergroup of Amoebozoa. This results are crucial for better understanding of the Amoebozoa taxa of both biomedical and evolutionary importance.

Keywords: Amoebozoa; MIA complex; Mitochondria; Mitosomes; OXA complex; PAM complex; Protein import; TIM22 complex; TIM23 complex; small Tims.

```
LOCUS       MF084995                570 bp    mRNA    linear   INV 08-JAN-2018
DEFINITION  Acanthamoeba castellanii Pam18 isoform A mRNA, complete cds.
ACCESSION   MF084995
VERSION     MF084995.1
KEYWORDS    .
SOURCE      Acanthamoeba castellanii
  ORGANISM  Acanthamoeba castellanii
            Eukaryota; Amoebozoa; Discosea; Longamoebia; Centramoebida;
            Acanthamoebidae; Acanthamoeba.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Wojtkowska,M., Buczek,D., Suzuki,Y., Shabardina,V., Makalowski,W.
            and Kmita,H.
  TITLE     The emerging picture of the mitochondrial protein import complexes
            of Amoebozoa supergroup
  JOURNAL   BMC Genomics 18 (1), 997 (2017)
   PUBMED   29284403
  REMARK    Publication Status: Online-Only
REFERENCE   2  (bases 1 to 570)
  AUTHORS   Buczek,D., Wojtkowska,M., Suzuki,Y., Makalowski,W. and Kmita,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2017) Adam Mickiewicz University, Institute of
            Molecular Biology and Biotechnology, Umultowska 89, Poznan 61-614,
            Poland
COMMENT     ##Assembly-Data-START##
            Assembly Method       :: Trinity De Novo Assembly v. 2012.10.05
            Sequencing Technology :: Illumina
            ##Assembly-Data-END##
FEATURES             Location/Qualifiers
     source          1..570
                     /organism="Acanthamoeba castellanii"
                     /mol_type="mRNA"
                     /db_xref="taxon:5755"
     CDS             113..406
                     /note="Pam18-A"
                     /codon_start=1
                     /product="Pam18 isoform A"
                     /protein_id="AUI80419.1"
                     /translation="MNAYKHFKAGNLTLPKGMVPKGPSRMKSYYTGGFESEMTRAEAA
                     LILSVRQGASKEKIKMAHRRIMLANHPDNGGSDYVASKVNEAKDLLLKDLGDD"
ORIGIN
        1 ccggaacttg catctgcgac catctgccgc tgtatcgatc cgggagcaat tggctactct
       61 tcttgtggcg ggtgtggtga ttgctggtgt ggccatcggt gggcgcgtgg ctatgaacgc
      121 ctacaagcac ttcaaggccg gcaacttgac cctccccaag ggcatggtgc ccaagggtcc
      181 atcgaggatg aagtcctact acacgggcgg ctttgagtcg gagatgaccc gcgccgaagc
      241 cgctctcatc ctcagtgtcc gacaaggcgc ctcgaaggag aagatcaaga tggcccacag
      301 gcggatcatg ttggcgaacc atcccgacaa tggaggcagc gactacgtgg cgtcgaaggt
      361 gaacgaagcc aaggacctgc tgctcaagga tctcggcgac gactgaggcc cctcttccaa
      421 aataacgaga gcaccaaaag gaccaccccc accaaacaca gacacacacc gagacaccac
      481 acaccgccgc cgccgccgcc accagagaga tcggggctgt acagagtact acaccaccat
      541 atatgcacac ccatattacc aacaaaaaaa
//
```

Pick Primers

Highlight Sequence Features

Find in this Sequence

Related information

Protein

PubMed

Taxonomy

Full text in PMC

PubMed (Weighted)

Recent activity

Turn Off Clear

Acanthamoeba castellanii Pam18 isoform A mRNA, complete cds — Nucleotide

Mus musculus muscle nicotinic acetylcholine receptor mRNA, partia — Nucleotide

makalowski AND cds (75473) — Nucleotide

Human Alu-Sb2 repeat, clone HALUSB11 — Nucleotide

makalowski (198935) — Nucleotide

See more...

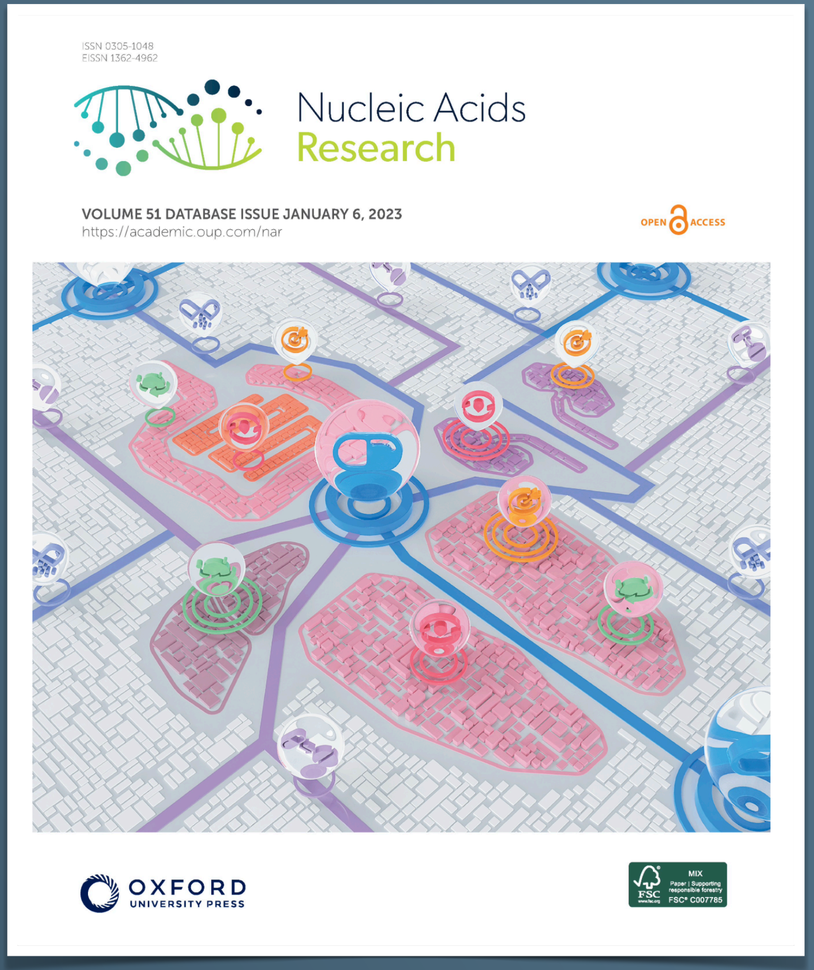
```
113..406
/note="Pam18-A"
/codon_start=1
/product="Pam18 isoform A"
/protein_id="AUI80419.1"
/translation="MNAYKHFKAGNLTLPKGMVPKGPSRMKSYYTGGFESEMTRAEAA
LILSVRQGASKEKIKMAHRRIMLANHPDNGGSDYVASKVNEAKDLLLKDLGDD"
```

CDS Feature 1 of 1 MF084995 : 1 segment

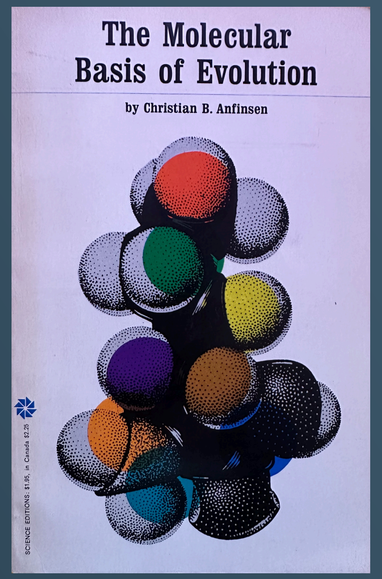
Details Display: FASTA GenBank Help

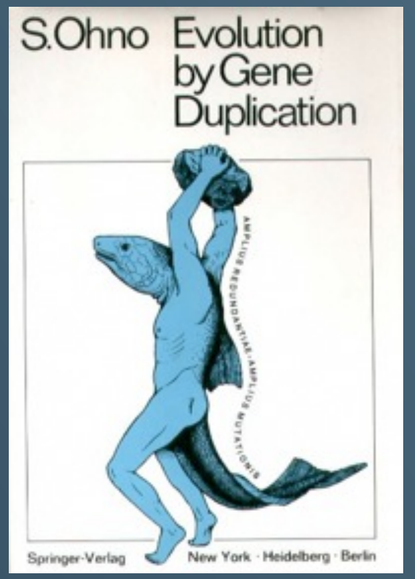
SECONDARY (SPECIALIZED) DATABASES

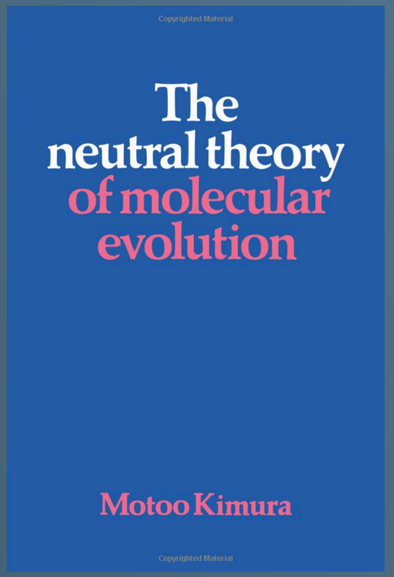
- Boom of biological databases
- Every year first issue of Nucleic Acids Research is dedicated to biological databases
  - https://academic.oup.com/nar/issue/51/D1
  - this year's database issue includes 1764 databases
  - the first collection published in 1993 contained description of 24 databases



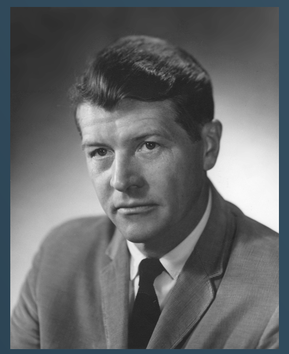
EVOLUTIONARY BASIS OF BIOINFORMATICS



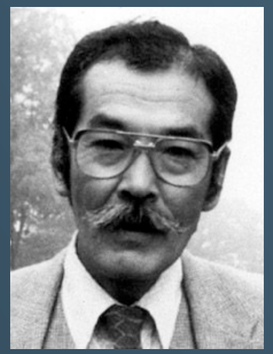




1959

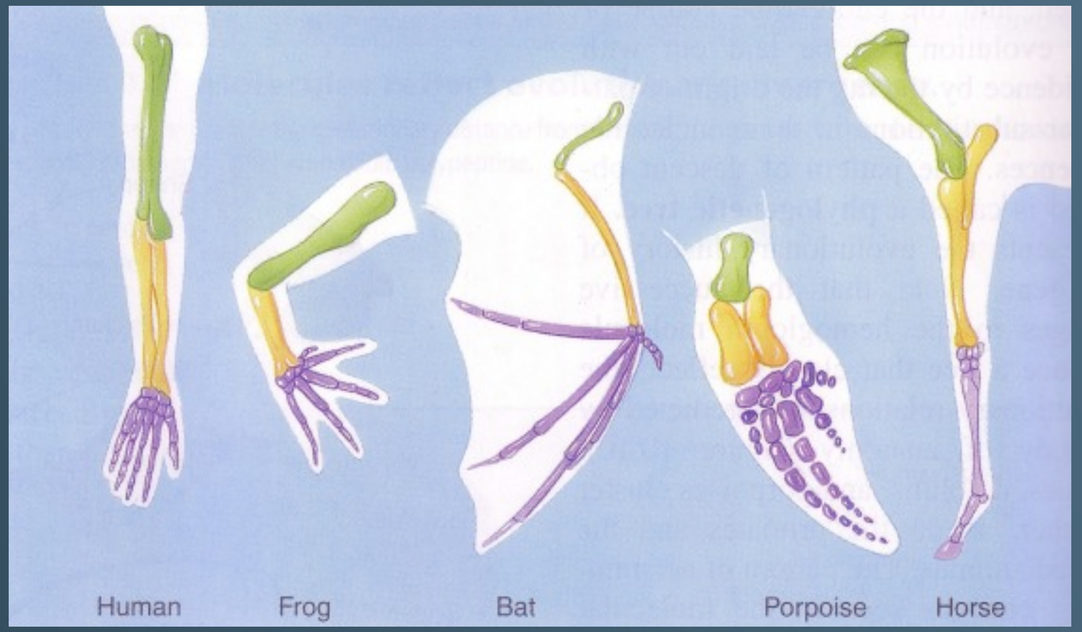


1970



1983

HOMOLOGS



Two anatomical structures or behavioral traits within different organisms which originated from a structure or trait of their common ancestral organism. The structures or traits in their current forms may not necessarily perform the same functions in each organism, nor perform the functions it did in the common ancestor. An example: the wing of a bat, the fin of a whale and the arm of a man are homologous structures.

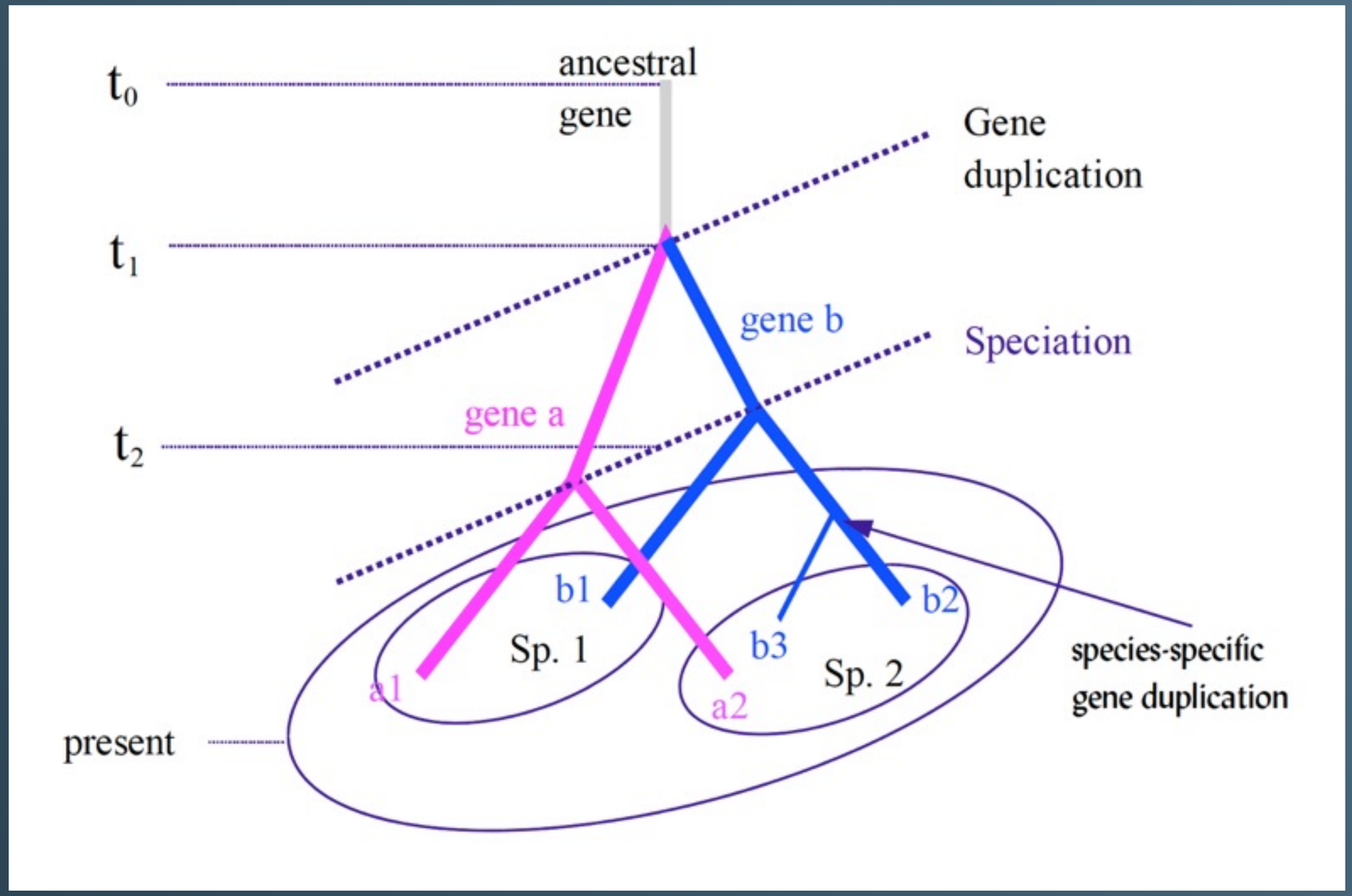
http://www.everythingbio.com/glos/definition.php?ID=3385

HOMOLOGS AT THE MOLECULAR LEVEL

```
cow       ATG---ACTAACATTCGAAAGTCCCACCCACTAATAAAAATTGTAAAC
sheep     ATG---ATCAACATCCGAAAAACCCACCCACTAATAAAAATTGTAAAC
goat      ATG---ACCAACATCCGAAAGACCCACCCATTAATAAAAATTGTAAAC
horse     ATG---ACAAACATCCGGAAATCTCACCCACTAATTAAAATCATCAAT
donkey    ATG---ACAAACATCCGAAAATCCCACCCGCTAATTAAAATCATCAAT
ostrich   ATGGCCCCCAACATTCGAAAATCGCACCCCCTGCTCAAAATTATCAAC
emu       ATGGCCCCTAACATCCGAAAATCCCACCCTCTACTCAAAATCATCAAC
turkey    ATGGCACCCAATATCCGAAAATCACACCCCCTATTAAAAACAATCAAC
```

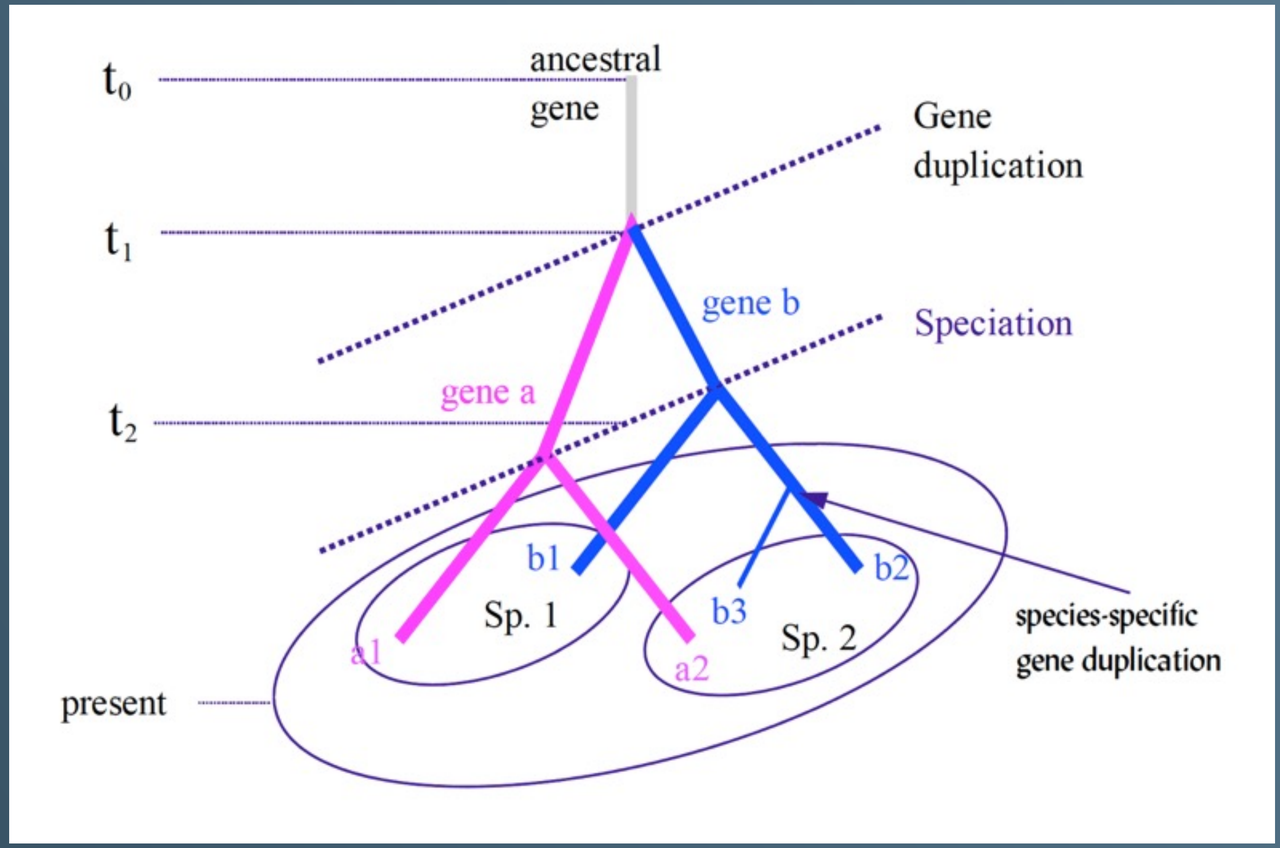
Two sequences that share common ancestry. Significant sequence similarity usually suggests homology, however sequence similarity may occur also by chance and some homologous sequences may diverge beyond detectable similarity.

EVOLUTIONARY BASIS OF BIOINFORMATICS



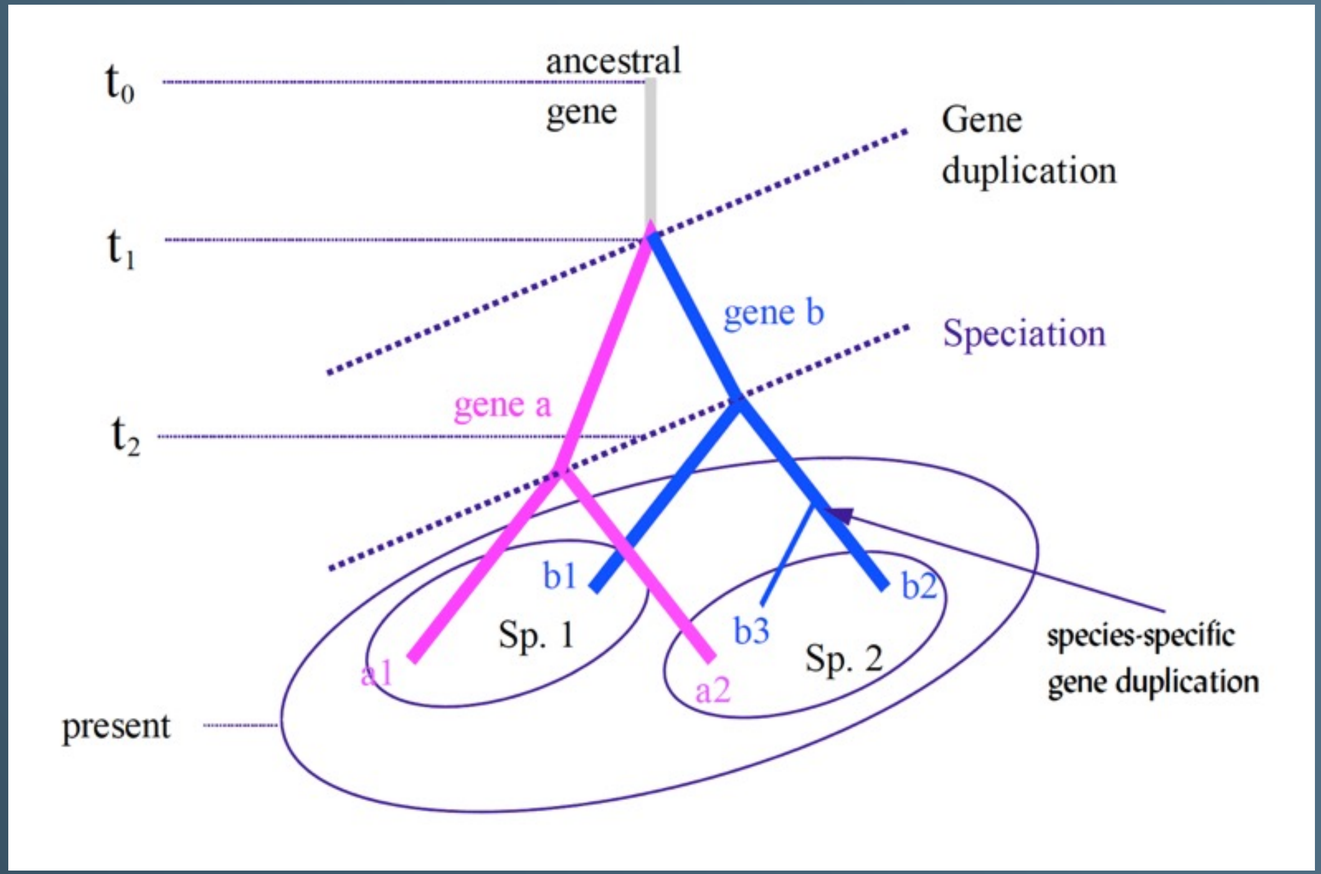
HOMOLOGS: ORTHOLOGS AND PARALOGS

ORTHOLOGS. Genes or sequences that result from a speciation event followed by a sequence divergence. Such genes cannot exist side by side in the same genome. The last common ancestor of two orthologous sequences existed just before speciation event.

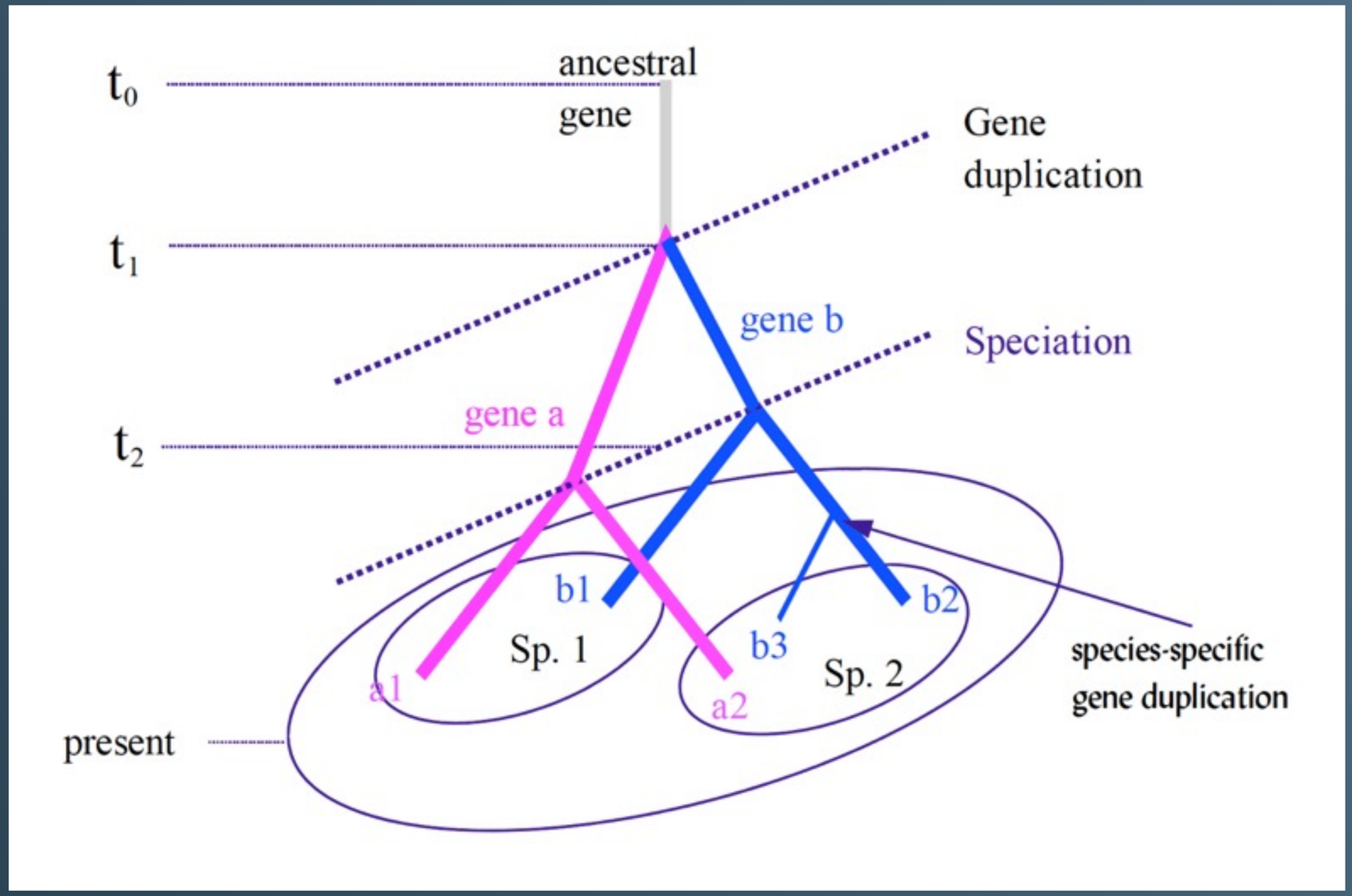


HOMOLOGS: ORTHOLOGS AND PARALOGS

PARALOGS. Genes or sequences that resulted from duplication of genetic material followed by a sequence divergence. Such genes may descend and diverge while existing side by side in the same genome. If speciation occurs after gene duplication, then two paralogous genes may exist in two different genomes. The last common ancestor of two paralogous sequences existed just before duplication event.

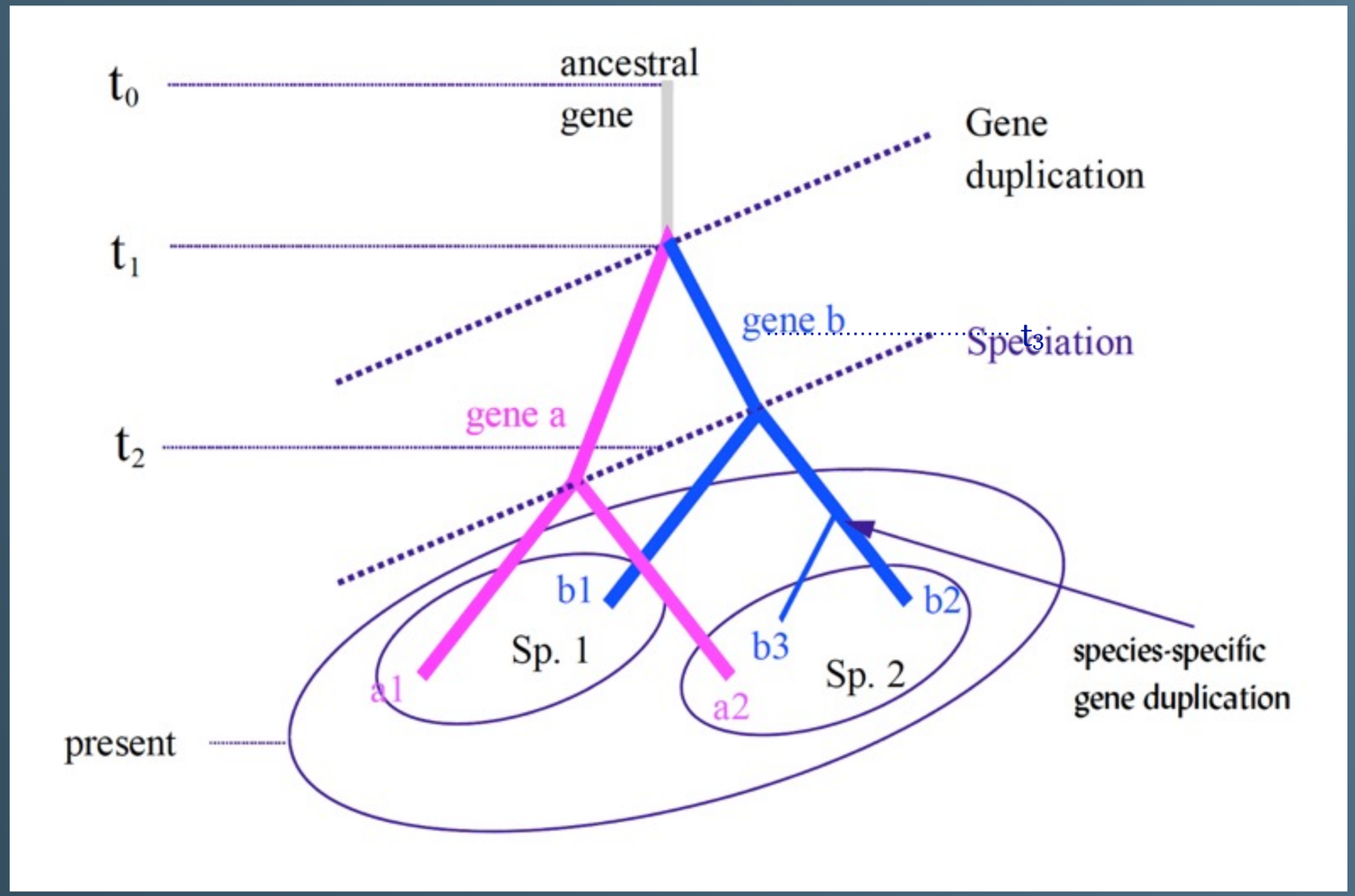


EVOLUTIONARY BASIS OF BIOINFORMATICS

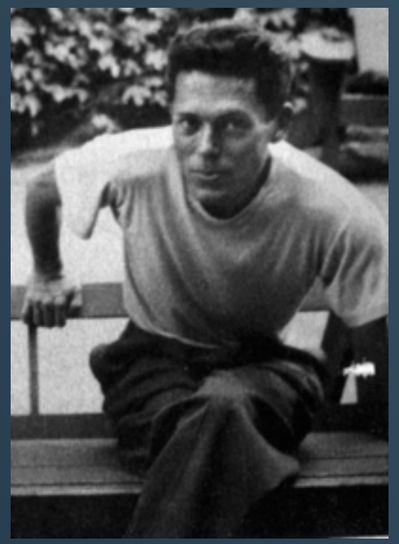


HOMOLOGS: ORTHOLOGS AND PARALOGS

Compared Genes	Relation	Time of the last comm. ancestor	Evolutionary event at the time of last common ancestor	Presence in the same species
A - B	paralogy	t_1	gene duplication	yes
A1 - A2	orthology	t_2	speciation	no
A1 - B1	paralogy	t_1	gene duplication	yes
A1 - B2	paralogy	t_1	gene duplication	no
A1 - B3	paralogy	t_1	gene duplication	no
A2 - A1	orthology	t_2	speciation	no
A2 - B1	paralogy	t_1	gene duplication	no
A2 - B2	paralogy	t_1	gene duplication	yes
A2 - B3	paralogy	t_1	gene duplication	yes
B1 - A1	paralogy	t_1	gene duplication	yes
B1 - A2	paralogy	t_1	gene duplication	no
B1 - B2	orthology	t_2	speciation	no
B1 - B3	orthology	t_2	speciation	no
B2 - A1	paralogy	t_1	gene duplication	no
B2 - A2	paralogy	t_1	gene duplication	yes
B2 - B1	orthology	t_2	speciation	no
B2 - B3	paralogy	t_3	gene duplication	yes
B3 - A1	paralogy	t_1	gene duplication	yes
B3 - A2	paralogy	t_1	gene duplication	no
B3 - B1	orthology	t_2	speciation	no
B3 - B2	paralogy	t_3	gene duplication	yes

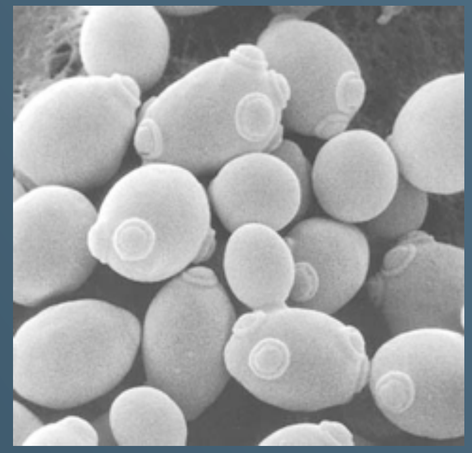


COMPARATIVE GENOMICS



What is true for *E. coli* is also true for elephant.
J. Monod, c. 1961

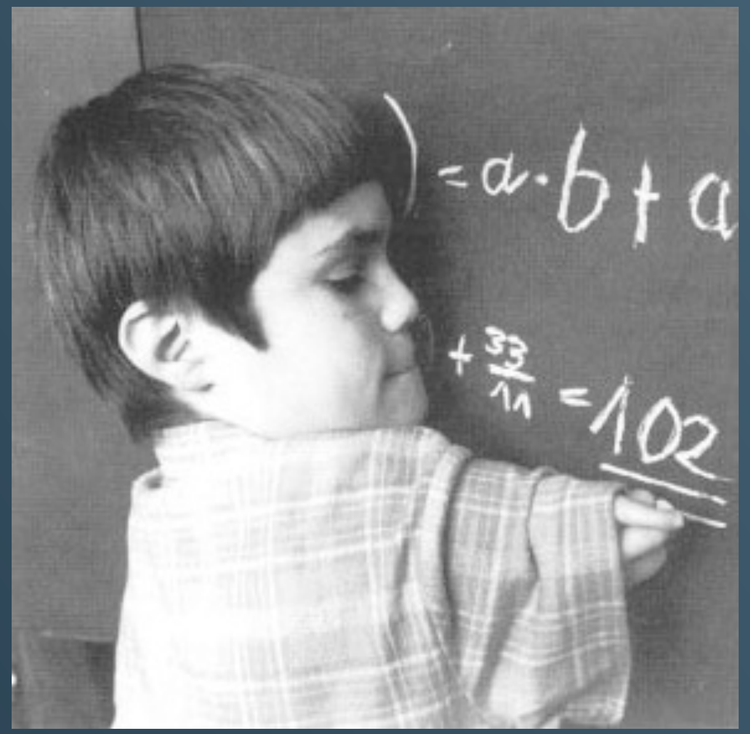
COMPARATIVE GENOMICS



What is true for yeast is also true for human.
D. Botstein, 1988

However...

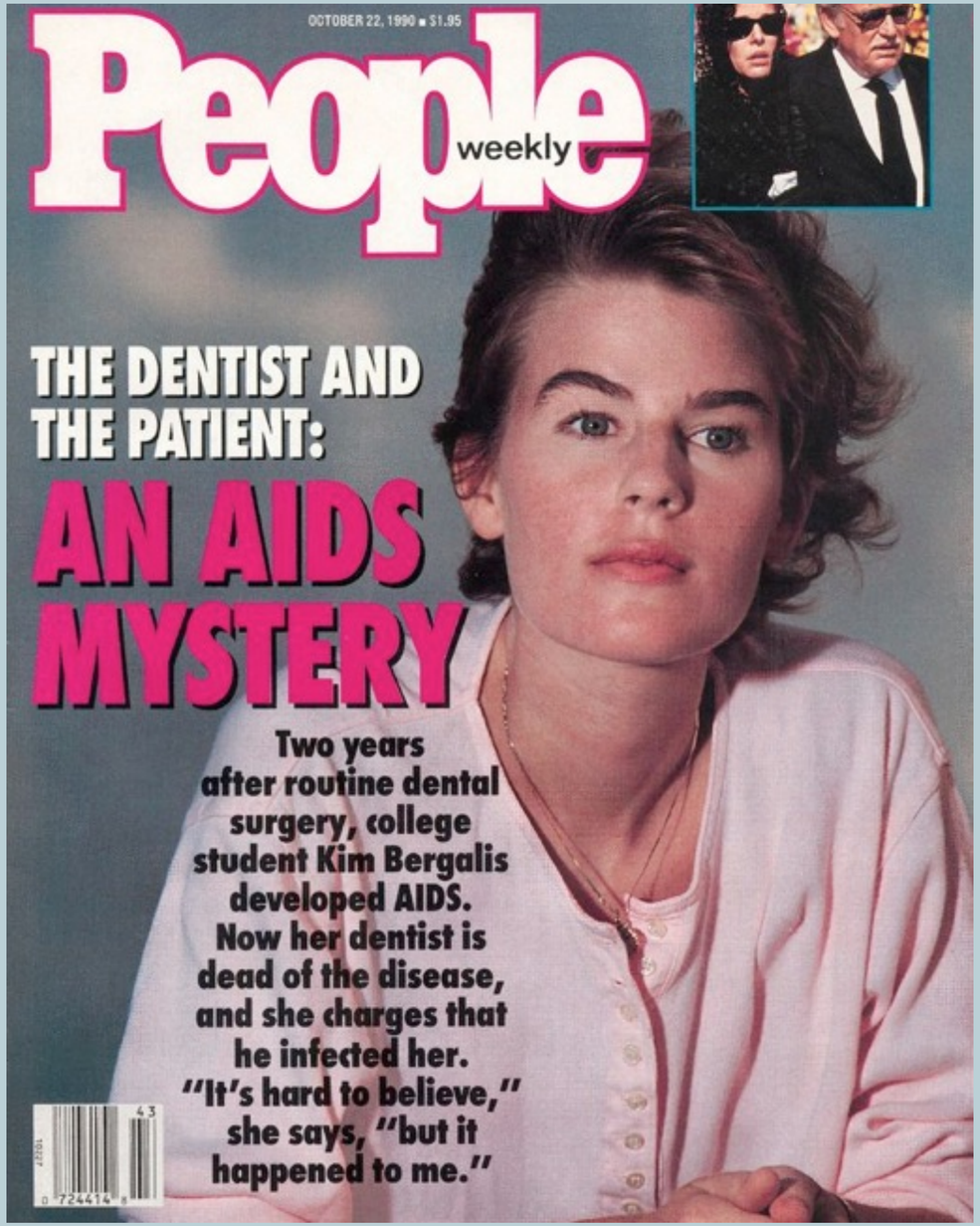
COMPARATIVE GENOMICS



10,000 victims of thalidomide

What is true for mouse is not necessarily true for human…


Is Bioinformatics Useful?



Did the Florida Dentist infect his patients with HIV?

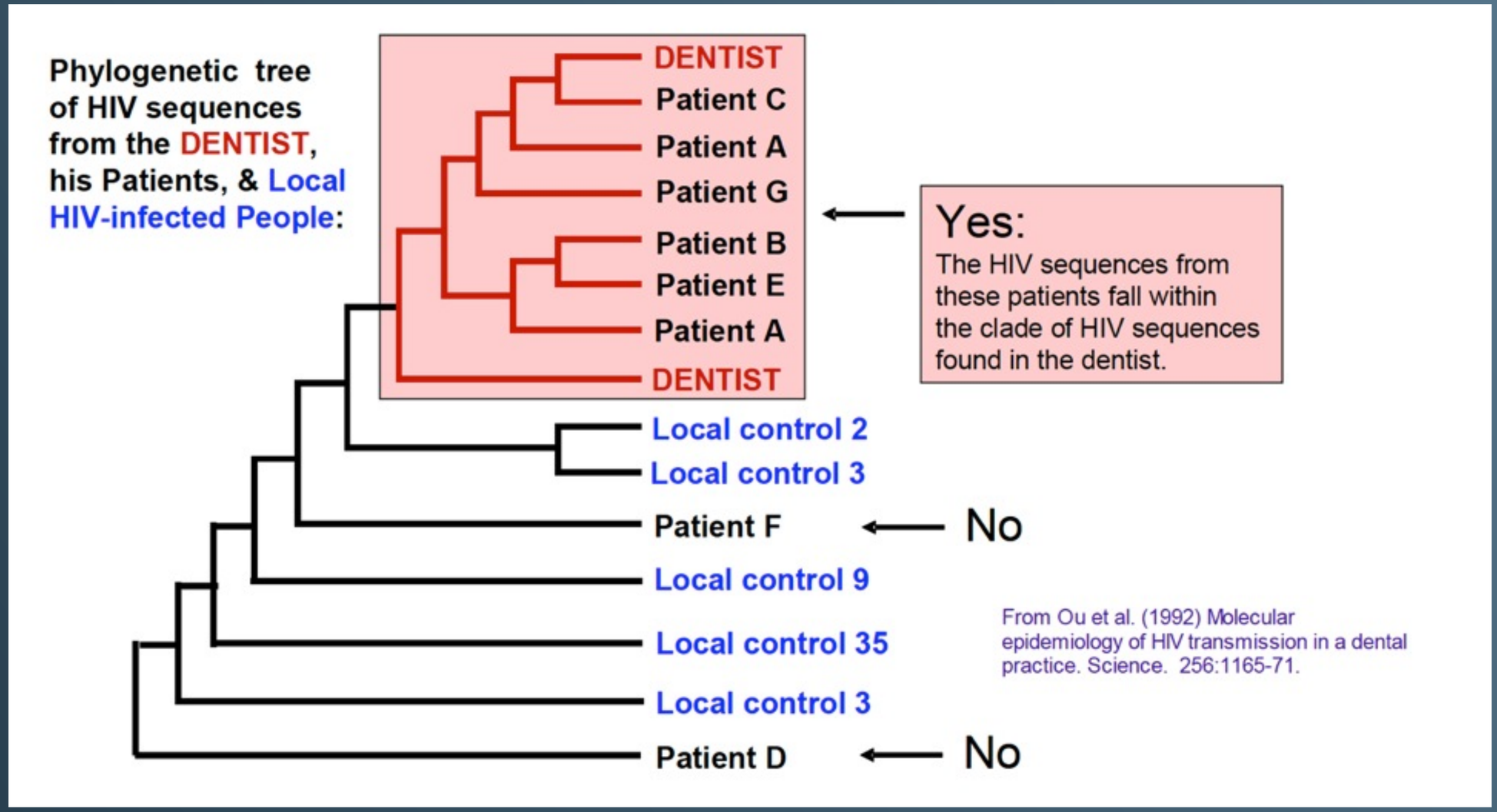
Kimberly Bergalis

(1968-1991)

David J. Acer

(1940-1990)

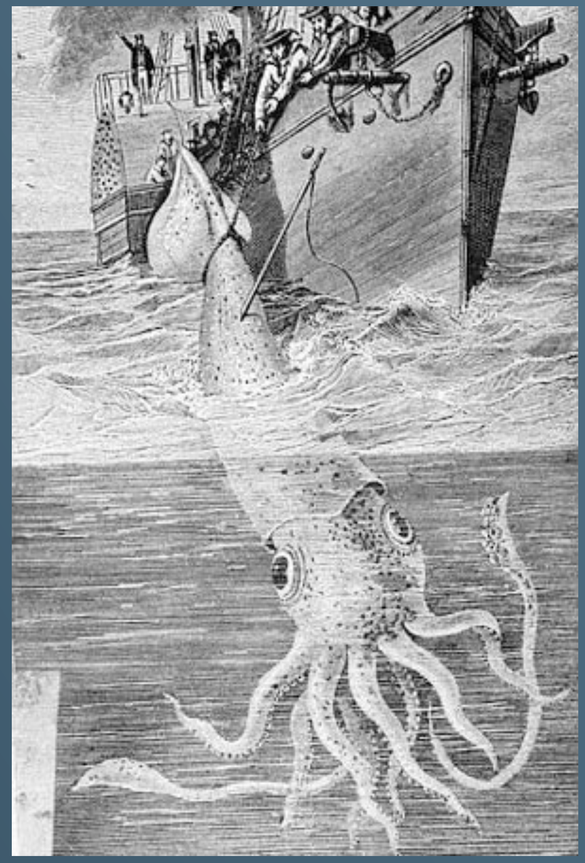
DID THE FLORIDA DENTIST INFECT HIS PATIENTS WITH HIV?



THE MYSTERY OF THE CHILEAN BLOB

THE MYSTERY OF THE CHILEAN BLOB

>Chilean_Blob
TAATACTAACTATATCCCTACTCTCCATTCTCATCGGGGGTT
GAGGAGGACTAAACCAGACTCAACTCCGAAAAATTATAGCTT
ACTCATCAATCGCCCACATAGGATGAATAACCACAATCCTAC
CCTACAATACAACCATAACCCTACTAAACCTACTAATCTATG
TCACAATAACCTTCACCATATTCATACTATTTATCCAAAACT
CAACCACAACCACACTATCTCTGTCCCAGACATGAAACAAAA
CACCCATTACCACAACCCTTACCATACTTACCCTACTTTCCA
TAGGGGGCCTCCCACCACTCTCGGGCTTTATCCCCAAATGAA
TAATTATTCAAGAACTAACAAAAAACGAAACCCTCATCATAC
CAACCTTCATAGCCACCACAGCATTACTCAACCTCTACTTCT
ATATACGCCTCACCTACTCAACAGCACTAACCCTATTCCCCT
CCACAAATAACATAAAAATAAAATGACAATTCTACCCCACAA
AACGAATAACCCTCCTGCCAACAGCAATTGTAATATCAACAA
TACTCCTACCCCTTACACCAATACTCTCCACCCTATTATAG



THE MYSTERY OF THE CHILEAN BLOB

Lineage Report

Cetacea [whales & dolphins]				
. Odontoceti [whales & dolphins]				
. . Physeteridae [whales & dolphins]				
. . . Physeter catodon	1085	3 hits	[whales & dolphins]	Physeter catodon NADH dehydrogenase subunit 2 (nad2) gene,
. . . Kogia breviceps	638	1 hit	[whales & dolphins]	Kogia breviceps complete mitochondrial genome
. . Orcaella brevirostris	593	1 hit	[whales & dolphins]	Orcaella brevirostris isolate 97 mitochondrion, complete ge
. . Grampus griseus	593	1 hit	[whales & dolphins]	Grampus griseus mitochondrion, complete genome
. . Feresa attenuata	592	2 hits	[whales & dolphins]	Feresa attenuata isolate 36 mitochondrion, complete genome
. . Tursiops truncatus (bottle-nosed dolphin)	592	1 hit	[whales & dolphins]	Tursiops truncatus mitochondrion, complete genome
. . Globicephala melas	586	3 hits	[whales & dolphins]	Globicephala melas isolate GlomelG42 mitochondrion, partial
. . Peponocephala electra	580	2 hits	[whales & dolphins]	Peponocephala electra isolate M6 mitochondrion, complete ge
. . Globicephala macrorhynchus	580	4 hits	[whales & dolphins]	Globicephala macrorhynchus isolate Glomac65 mitochondrion,
. . Pseudorca crassidens	577	3 hits	[whales & dolphins]	Pseudorca crassidens mitochondrion, complete genome
. . Orcinus orca (Orca)	569	54 hits	[whales & dolphins]	Orcinus orca isolate ENPTGA2 mitochondrion, complete genome
. . Sotalia fluviatilis	569	2 hits	[whales & dolphins]	Sotalia fluviatilis haplotype 10 NADH dehydrogenase subunit
. . Platanista minor	569	1 hit	[whales & dolphins]	Platanista minor complete mitochondrial genome
. . Steno bredanensis	566	2 hits	[whales & dolphins]	Steno bredanensis isolate StebreS9 mitochondrion, partial g
. Megaptera novaeangliae	636	5 hits	[whales & dolphins]	Megaptera novaeangliae voucher GOM9049 NADH dehydrogenase s
. Balaenoptera bonaerensis	630	1 hit	[whales & dolphins]	Balaenoptera bonaerensis mitochondrial DNA, complete genome
. Eubalaena japonica	619	1 hit	[whales & dolphins]	Eubalaena japonica mitochondrial DNA, complete genome
. Balaenoptera brydei	614	2 hits	[whales & dolphins]	Balaenopte
. Balaena mysticetus (Greenland right whale)	614	2 hits	[whales & dolphins]	Balaena my
. Balaenoptera musculus	614	1 hit	[whales & dolphins]	Balaenopte
. Balaenoptera edeni	603	1 hit	[whales & dolphins]	Balaenopte
. Balaenoptera omurai	603	2 hits	[whales & dolphins]	Balaenopte
. Eschrichtius robustus (California gray whale)	603	2 hits	[whales & dolphins]	Eschrichti
. Balaenoptera borealis	597	1 hit	[whales & dolphins]	Balaenopte
. Caperea marginata	580	1 hit	[whales & dolphins]	Caperea ma
. Balaenoptera physalus (finback whale)	569	1 hit	[whales & dolphins]	Balaenopte

THE MYSTERY OF THE CHILEAN BLOB

```
>emb|AJ277029.2|  Physeter macrocephalus mitochondrial genome
Length=16428

 Score = 1074 bits (581),  Expect = 0.0
 Identities = 585/587 (99%), Gaps = 0/587 (0%)
 Strand=Plus/Plus

Query  1     TAATACTAACTATATCCCTACTCTCCATTCTCATCGGGGGTTGAGGAGGACTAAACCAGA  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4400  TAATACTAACTATATCCCTACTCTCCATTCTCATCGGGGGTTGAGGAGGACTAAACCAGA  4459

Query  61    CTCAACTCCGAAAAATTATAGCTTACTCATCAATCGCCCACATAGGATGAATAACCACAA  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4460  CTCAACTCCGAAAAATTATAGCTTACTCATCAATCGCCCACATAGGATGAATAACCACAA  4519

Query  121   TCCTACCCTACAATACAACCATAACCCTACTAAACCTACTAATCTATGTCACAATAACCT  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4520  TCCTACCCTACAATACAACCATAACCCTACTAAACCTACTAATCTATGTCACAATAACCT  4579

Query  181   TCACCATATTCATACTATTTATCCAAAACTCAACCACAACCACACTATCTCTGTCCCAGA  240
             |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4580  TCACCATATTCACACTATTTATCCAAAACTCAACCACAACCACACTATCTCTGTCCCAGA  4639

Query  241   CATGAAACAAAACACCCATTACCACAACCCTTACCATACTTACCCTACTTTCCATAGGGG  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4640  CATGAAACAAAACACCCATTACCACAACCCTTACCATACTTACCCTACTTTCCATAGGGG  4699

Query  301   GCCTCCCACCACTCTCGGGCTTTATCCCCAAATGAATAATTATTCAAGAACTAACAAAAA  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4700  GCCTCCCACCACTCTCGGGCTTTATCCCCAAATGAATAATTATTCAAGAACTAACAAAAA  4759

Query  361   ACGAAACCCTCATCATACCAACCTTCATAGCCACCACAGCATTACTCAACCTCTACTTCT  420
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4760  ACGAAGCCCTCATCATACCAACCTTCATAGCCACCACAGCATTACTCAACCTCTACTTCT  4819

Query  421   ATATACGCCTCACCTACTCAACAGCACTAACCCTATTCCCCTCCACAAATAACATAAAAA  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4820  ATATACGCCTCACCTACTCAACAGCACTAACCCTATTCCCCTCCACAAATAACATAAAAA  4879

Query  481   TAAAATGACAATTCTACCCCACAAAACGAATAACCCTCCTGCCAACAGCAATTGTAATAT  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4880  TAAAATGACAATTCTACCCCACAAAACGAATAACCCTCCTGCCAACAGCAATTGTAATAT  4939

Query  541   CAACAATACTCCTACCCCTTACACCAATACTCTCCACCCTATTATAG  587
             |||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4940  CAACAATACTCCTACCCCTTACACCAATACTCTCCACCCTATTATAG  4986
```


where
to
start

THE BEST IS TO START AT ONE OF CENTRAL REPOSITORIES

NCBI
National Center for Biotechnology Information

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

EMBL-EBI
The home for big data in biology

https://www.ebi.ac.uk

DDBJ
DNA Data Bank of Japan

https://www.ddbj.nig.ac.jp/index-e.html

NCBI Resources How To Sign in to NCBI

NCBI
National Center for Biotechnology Information

All Databases search term here, e.g. protein name Search

COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: https://www.coronavirus.gov .
Get the latest research from NIH: https://www.nih.gov/coronavirus.
Find NCBI SARS-CoV-2 literature, sequence, and clinical content: https://www.ncbi.nlm.nih.gov/sars-cov-2/.

- NCBI Home
- Resource List (A-Z)
- All Resources
- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis
- Taxonomy
- Training & Tutorials
- Variation

Welcome to NCBI

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

About the NCBI | Mission | Organization | NCBI News & Blog

Submit
Deposit data or manuscripts into NCBI databases

Download
Transfer NCBI data to your computer

Learn
Find help documents, attend a class or watch a tutorial

Develop
Use NCBI APIs and code libraries to build applications

Analyze
Identify an NCBI tool for your data analysis task

Research
Explore NCBI research and collaborative projects

Popular Resources

- PubMed
- Bookshelf
- PubMed Central
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein
- PubChem

NCBI News & Blog

IgBLAST 1.17 is now available with improved identification of productive V gene sequences
30 Oct 2020

A new release of IgBLAST (1.17), the

New feature in the dbGap submission portal: Automated study metadata
29 Oct 2020

dbGaP has recently released a new feature to simplify submissions and

Search NCBI "globin x" Search

Results found in 20 databases

Literature	
Bookshelf	0
MeSH	1
NLM Catalog	5
PubMed	32
PubMed Central	101

Genes	
Gene	179
GEO DataSets	436
GEO Profiles	224
HomoloGene	3
PopSet	2

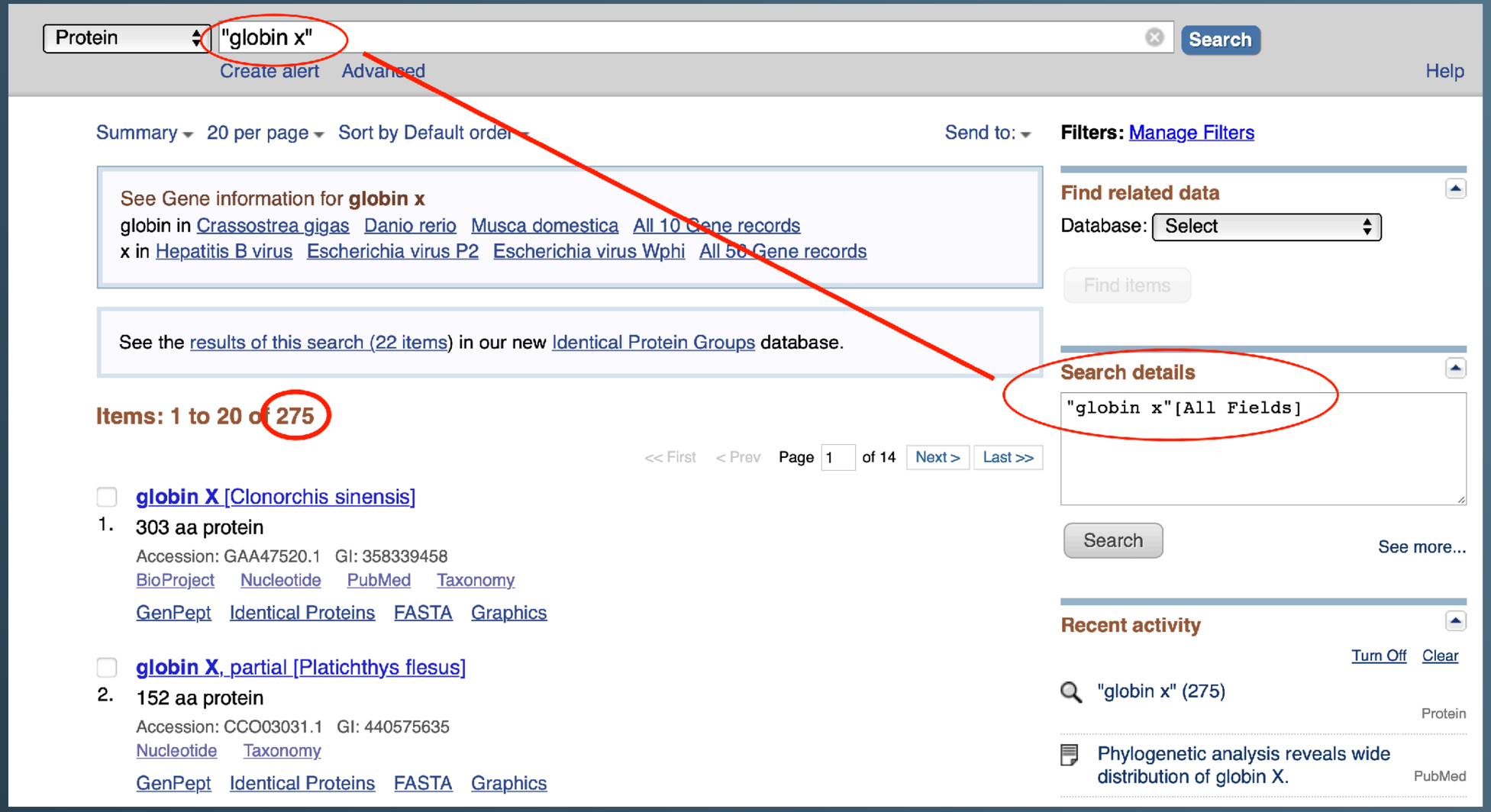
Proteins	
Conserved Domains	3
Identical Protein Groups	38
Protein	426
Protein Family Models	1
Structure	742

Genomes	
Assembly	7,566
BioCollections	0
BioProject	31
BioSample	0
Genome	0
Nucleotide	315
SRA	0
Taxonomy	0

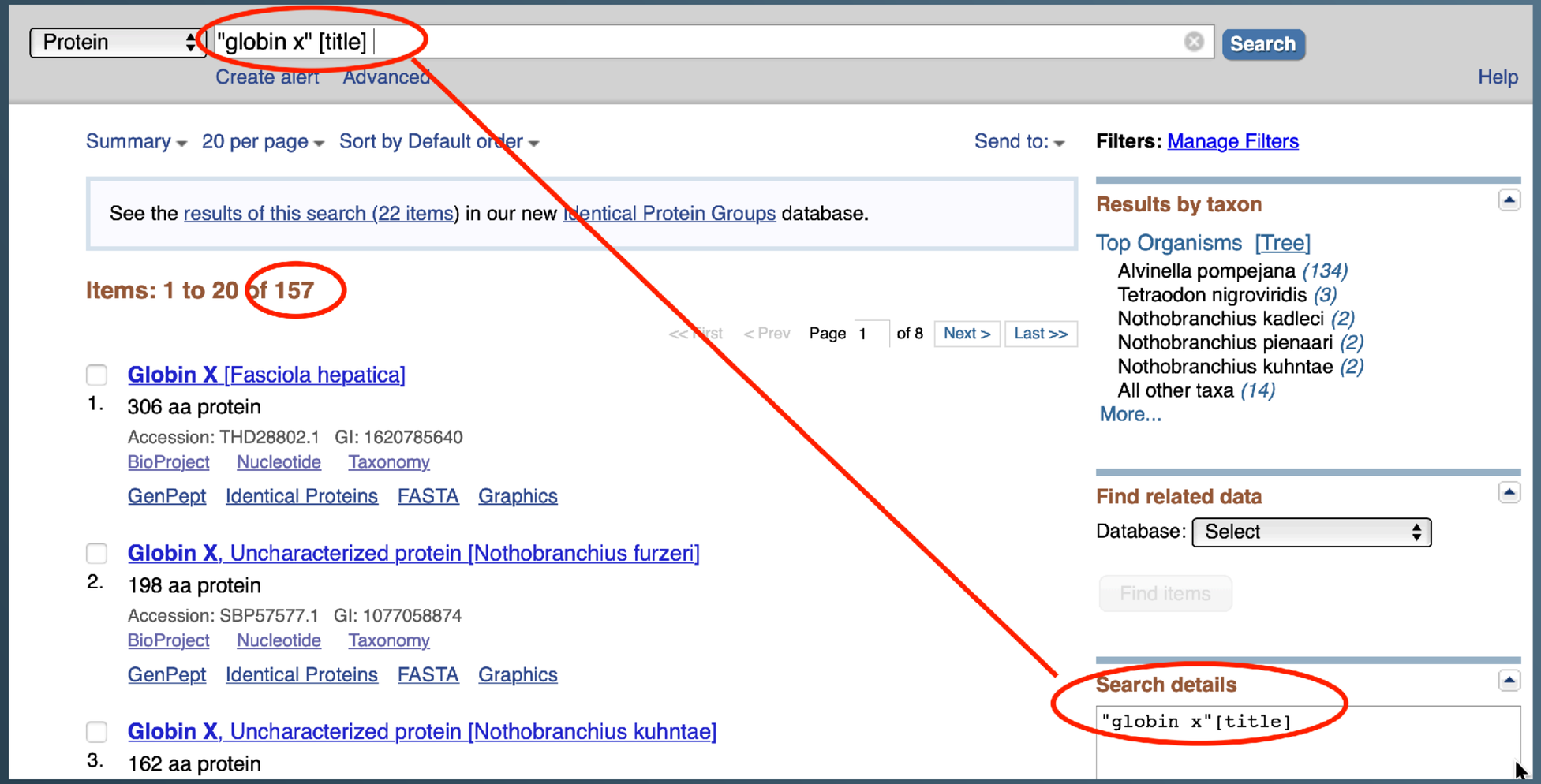
Clinical	
ClinicalTrials.gov	0
ClinVar	0
dbGaP	0
dbSNP	0
dbVar	35
GTR	55
MedGen	0
OMIM	91

PubChem	
BioAssays	0
Compounds	0
Pathways	0
Substances	0

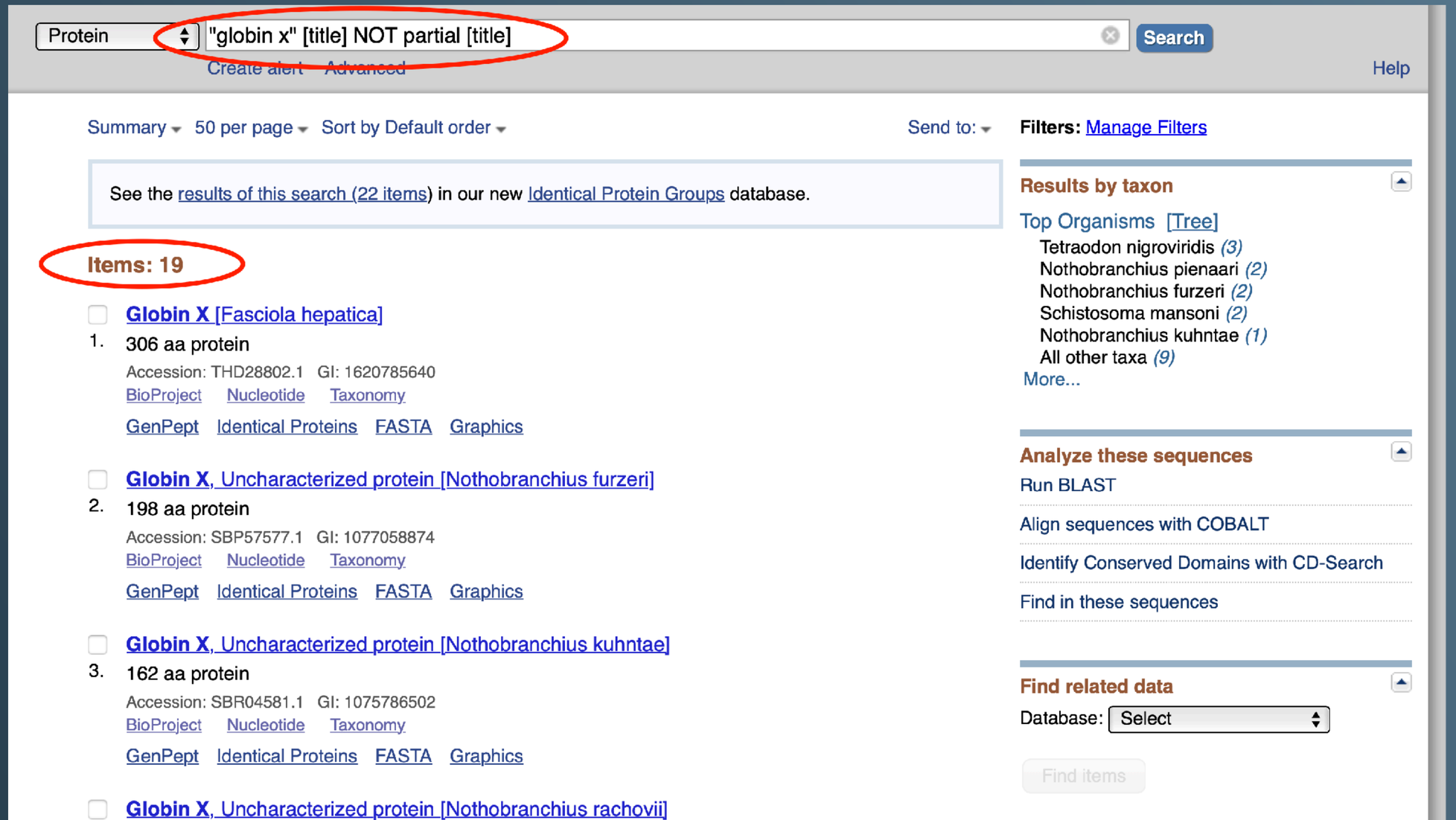
LET'S SEARCH FOR "GLOBIN X" SEQUENCES



WE CAN MAKE RESULTS MORE SPECIFIC



WE CAN MAKE RESULTS *EVEN* MORE SPECIFIC



THE COOL STUFF STARTS *NOW!*

Protein | Protein | "globin x" [title] NOT partial [title] AND vertebrata [organism] | Search

Create alert | Advanced | Help

COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: https://www.coronavirus.gov .
Get the latest research from NIH: https://www.nih.gov/coronavirus.
Find NCBI SARS-CoV-2 literature, sequence, and clinical content: https://www.ncbi.nlm.nih.gov/sars-cov-2/.

Species
Animals (15)
Customize ...

Source databases
Customize ...

Sequence length
Custom range...

Molecular weight
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: | **Filters:** Manage Filters

See the results of this search (14 items) in our new Identical Protein Groups database.

Items: 15

1. **Globin X**, Uncharacterized protein [Nothobranchius furzeri]
198 aa protein
Accession: SBP57577.1 GI: 1077058874
BioProject Nucleotide Taxonomy
GenPept Identical Proteins FASTA Graphics

2. **Globin X**, Uncharacterized protein [Nothobranchius kuhntae]
162 aa protein
Accession: SBR04581.1 GI: 1075786502
BioProject Nucleotide Taxonomy
GenPept Identical Proteins FASTA Graphics

3. **Globin X**, Uncharacterized protein [Nothobranchius rachovii]
198 aa protein
Accession: SBR92617.1 GI: 1075753413
BioProject Nucleotide Taxonomy
GenPept Identical Proteins FASTA Graphics

Results by taxon

Top Organisms [Tree]
Tetraodon nigroviridis *(3)*
Nothobranchius pienaari *(2)*
Nothobranchius furzeri *(2)*
Nothobranchius kuhntae *(1)*
Nothobranchius rachovii *(1)*
All other taxa *(6)*
More...

Analyze these sequences

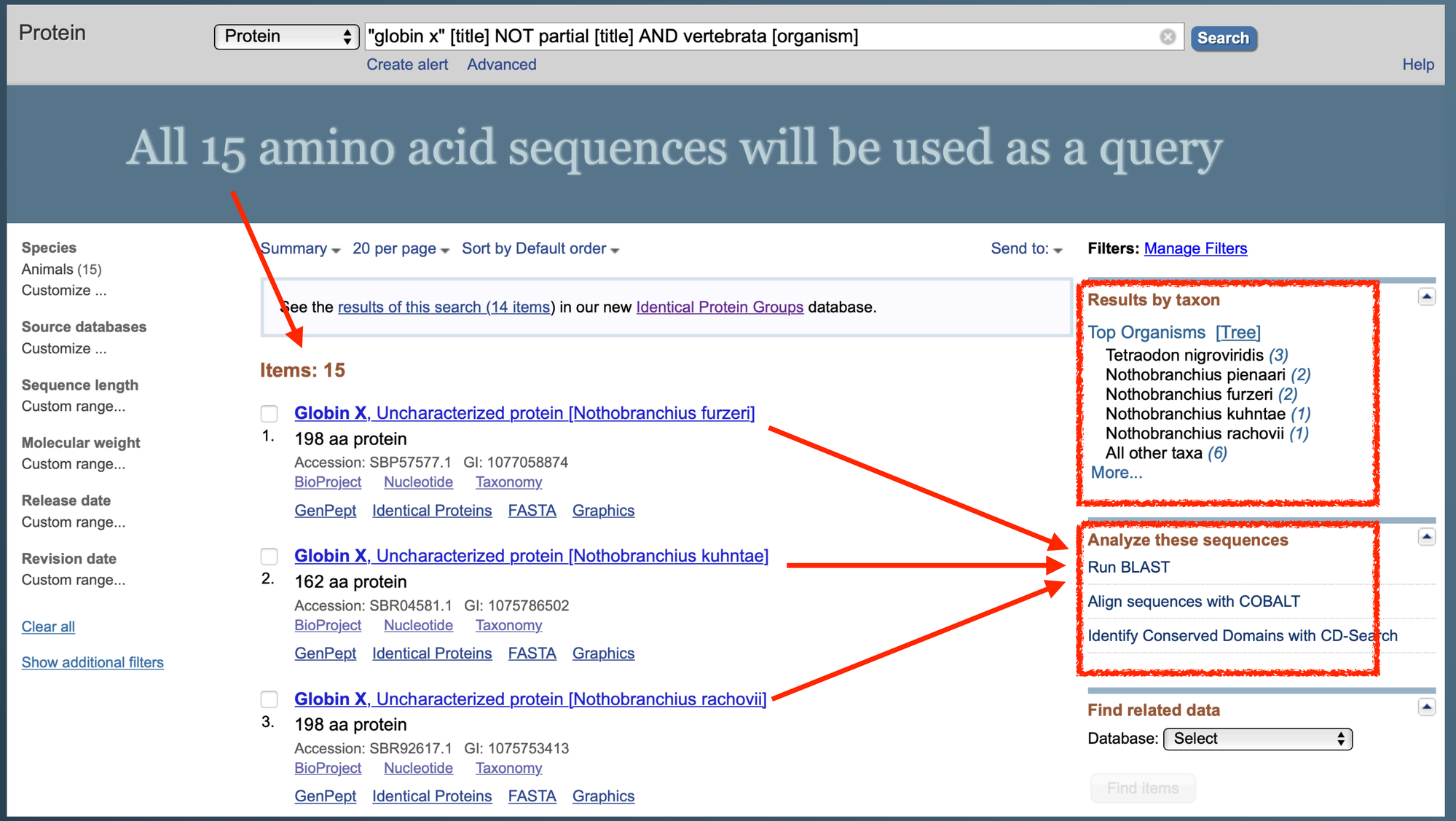
Run BLAST
Align sequences with COBALT
Identify Conserved Domains with CD-Search

Find related data

Database: Select

Find items

LET'S RUN BLAST



NIH U.S. National Library of Medicine | NCBI National Center for Biotechnology Information | Sign in to NCBI

BLAST® » blastp suite

Home | Recent Results | Saved Strategies | Help

Standard Protein BLAST

blastn | **blastp** | blastx | tblastn | tblastx

BLASTP programs search protein databases using a protein query. more...

Reset page | Bookmark

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) Clear

THD28802.1
SBP57577.1
SBR04581.1
SBR92617.1
SBQ48984.1

List of proteins from our search

Query subrange

From

To

Or, upload file Choose File no file selected

Job Title

Enter a descriptive title for your BLAST search

☐ **Align two or more sequences**

We are beta testing a New Results page

☑ **Click here if you would like to see your results in the new format. You can always switch back to the Traditional Results page.**

Choose Search Set

Database Non-redundant protein sequences (nr)

Organism Optional

Enter organism name or id--completions will be suggested ☐ exclude +

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Exclude Optional

☐ Models (XM/XP) ☐ Non-redundant RefSeq proteins (WP) ☐ Uncultured/environmental sample sequences

Database choice and results limitation options

Program Selection

Algorithm

- ○ Quick BLASTP (Accelerated protein-protein BLAST)
- ◉ blastp (protein-protein BLAST)
- ○ PSI-BLAST (Position-Specific Iterated BLAST)
- ○ PHI-BLAST (Pattern Hit Initiated BLAST)
- ○ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm

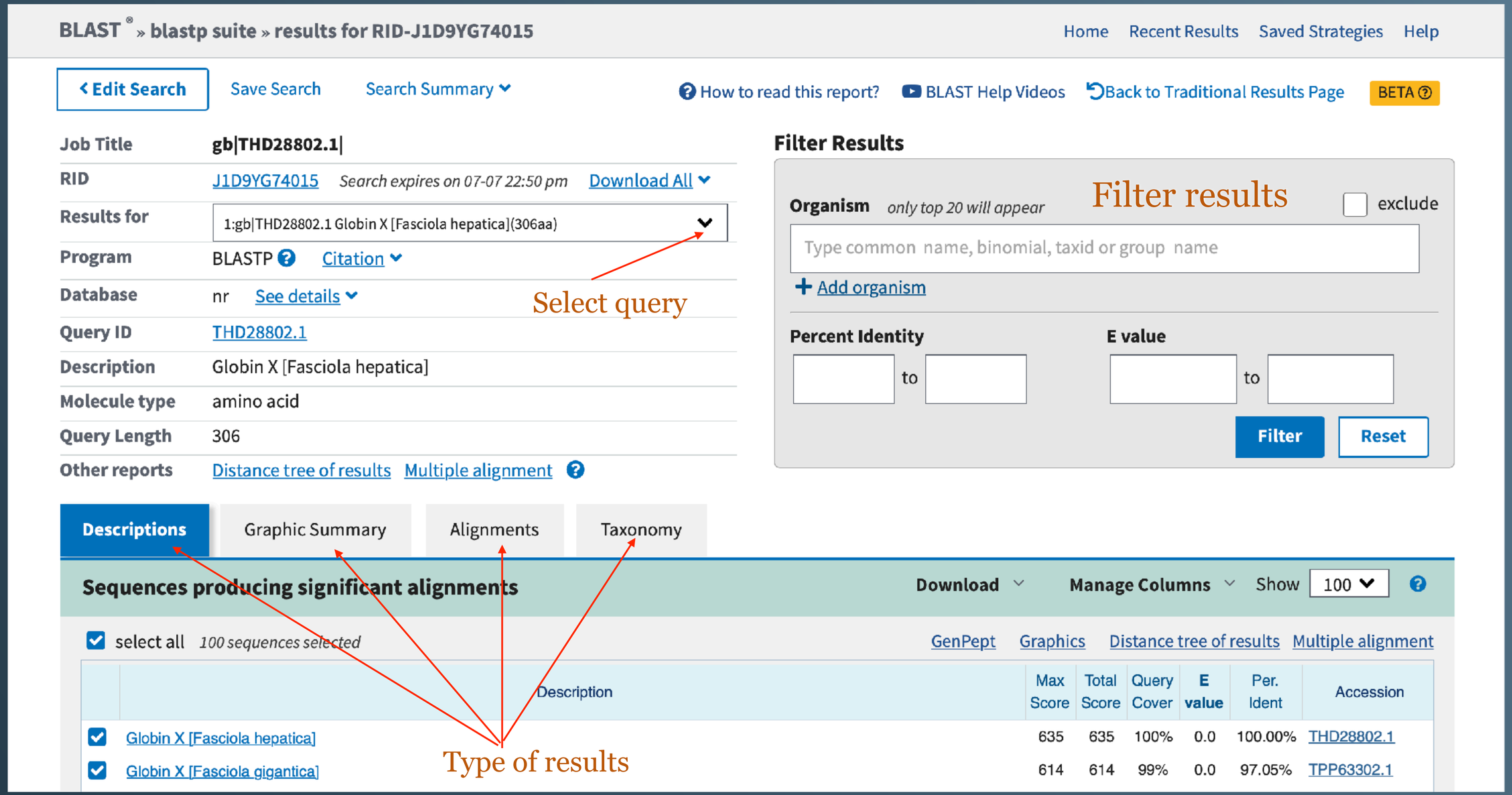
Different types (algorithms) of BLAST

BLAST Search database nr using Blastp (protein-protein BLAST)

☐ Show results in a new window

Klick here to start BLAST

BLAST RESULTS



Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download Manage Columns Show 100

select all *100 sequences selected* GenPept Graphics Distance tree of results Multiple alignment

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
Globin X [Fasciola hepatica]	635	635	100%	0.0	100.00%	THD28802.1
Globin X [Fasciola gigantica]	614	614	99%	0.0	97.05%	TPP63302.1
hypothetical protein CRM22_002376 [Opisthorchis felineus]	304	304	85%	1e-99	56.49%	TGZ71927.1
unnamed protein product [Echinostoma caproni]	297	297	69%	3e-98	68.20%	VDP67930.1
hypothetical protein T265_04650 [Opisthorchis viverrini]	300	300	93%	7e-98	52.08%	XP_009167705.1
hypothetical protein CSKR_5917s [Clonorchis sinensis]	298	298	87%	2e-97	54.78%	RJW73736.1
globin [Opisthorchis viverrini]	295	295	92%	3e-96	52.11%	OON21854.1
Leghemoglobin-1 isoform 1 [Schistosoma japonicum]	218	218	83%	2e-65	42.15%	TNN15233.1
SJCHGC09035 protein [Schistosoma japonicum]	204	204	58%	1e-61	50.00%	AAW24922.1
unnamed protein product [Schistosoma margrebowiei]	205	205	77%	3e-60	44.03%	VDP16629.1
uncharacterized protein DC041_0005585 [Schistosoma bovis]	201	201	77%	9e-59	43.22%	RTG83182.1
unnamed protein product [Schistosoma mattheei]	168	168	61%	1e-46	43.08%	VDP64096.1
PREDICTED: cytoglobin-2-like [Biomphalaria glabrata]	115	115	48%	3e-27	35.81%	XP_013084131.1
PREDICTED: cytoglobin-1-like [Callorhinchus milii]	112	112	54%	4e-26	35.71%	XP_007891388.1
PREDICTED: neuroglobin-like [Haplochromis burtoni]	111	111	52%	4e-26	37.89%	XP_005952972.2
PREDICTED: neuroglobin-like [Gekko japonicus]	112	112	48%	5e-26	37.58%	XP_015274271.1
neuroglobin isoform X2 [Oreochromis niloticus]	110	110	48%	9e-26	37.33%	XP_019201382.1
PREDICTED: neuroglobin-like [Neolamprologus brichardi]	110	110	48%	2e-25	37.33%	XP_006793470.1
PREDICTED: neuroglobin-like isoform X2 [Pundamilia nyererei]	110	110	48%	2e-25	37.33%	XP_005754405.1
neuroglobin-like isoform X1 [Erpetoichthys calabaricus]	110	110	48%	3e-25	36.49%	XP_028653231.1
GbX2 [Callorhinchus milii]	108	108	45%	3e-24	36.69%	AKU74647.1

Definition from the source database

Select number of results to see.

Warning: if you want to see more than 100 results, you need to specify it on the first screen under "Algorithm parameters"

Descriptions | **Graphic Summary** | Alignments | Taxonomy

hover to see the title *click to show alignments*

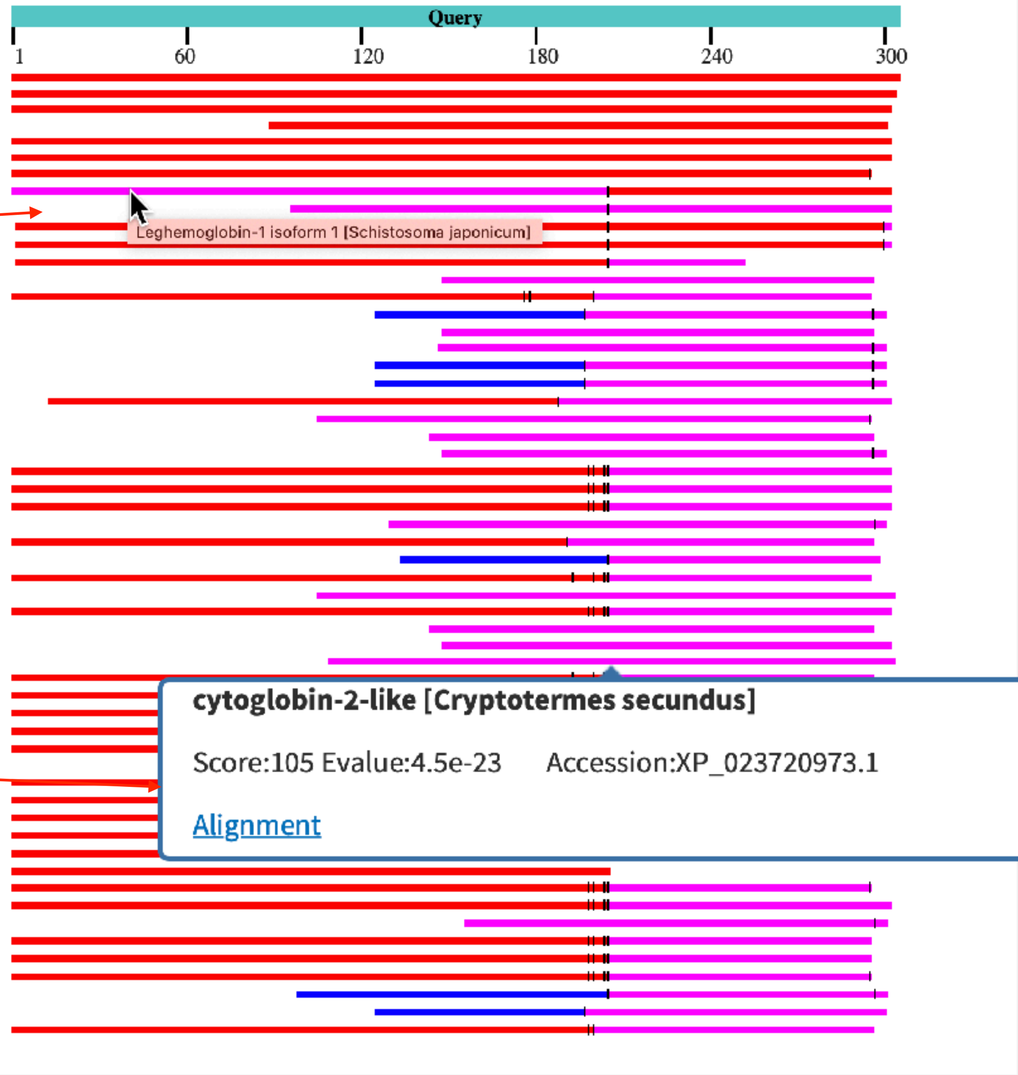
Alignment Scores < 40 | 40 - 50 | 50 - 80 | 80 - 200 | >= 200

Alignment score: higher is better 👍

100 sequences selected

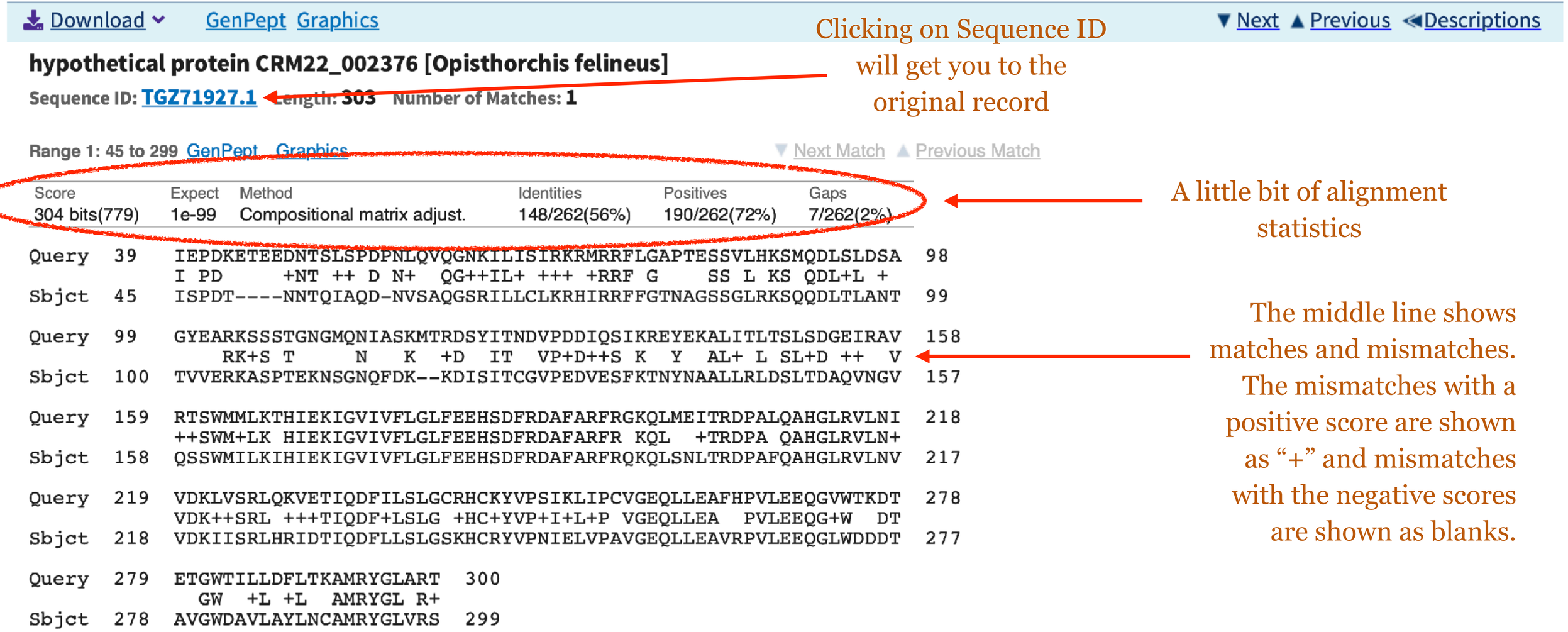
Positioning the cursor over a bar shows a definition line of the hit (a sequence from the database)

Clicking on a bar will display more info including link to the alignment



on this page in a new tab

BLAST RESULTS: ALIGNMENTS

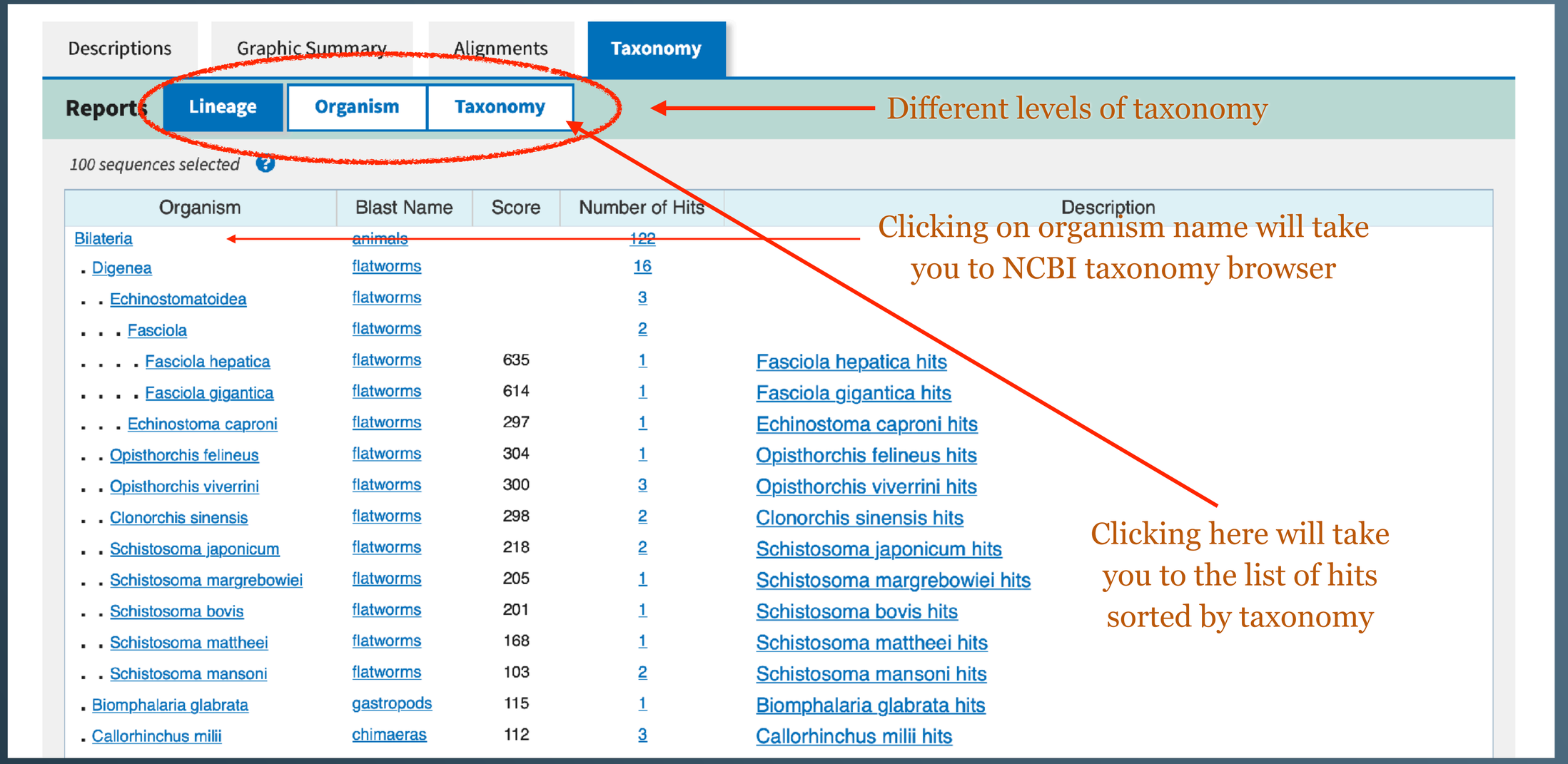


Clicking on Sequence ID will get you to the original record

A little bit of alignment statistics

The middle line shows matches and mismatches. The mismatches with a positive score are shown as "+" and mismatches with the negative scores are shown as blanks.

BLAST RESULTS: TAXONOMY



Organism	Blast Name	Score	Number of Hits	Description
Bilateria	animals		122	
. Digenea	flatworms		16	
. . Echinostomatoidea	flatworms		3	
. . . Fasciola	flatworms		2	
. . . . Fasciola hepatica	flatworms	635	1	Fasciola hepatica hits
. . . . Fasciola gigantica	flatworms	614	1	Fasciola gigantica hits
. . . Echinostoma caproni	flatworms	297	1	Echinostoma caproni hits
. . Opisthorchis felineus	flatworms	304	1	Opisthorchis felineus hits
. . Opisthorchis viverrini	flatworms	300	3	Opisthorchis viverrini hits
. . Clonorchis sinensis	flatworms	298	2	Clonorchis sinensis hits
. . Schistosoma japonicum	flatworms	218	2	Schistosoma japonicum hits
. . Schistosoma margrebowiei	flatworms	205	1	Schistosoma margrebowiei hits
. . Schistosoma bovis	flatworms	201	1	Schistosoma bovis hits
. . Schistosoma mattheei	flatworms	168	1	Schistosoma mattheei hits
. . Schistosoma mansoni	flatworms	103	2	Schistosoma mansoni hits
. Biomphalaria glabrata	gastropods	115	1	Biomphalaria glabrata hits
. Callorhinchus milii	chimaeras	112	3	Callorhinchus milii hits

BLAST RESULTS: TAXONOMY

Descriptions | Graphic Summary | Alignments | **Taxonomy**

Reports | **Lineage** | **Organism** | **Taxonomy**

100 sequences selected

Taxonomy	Number of hits	Number of Organisms	Description
⊟Teleostei	116	94	
. ⊟Clupeocephala	115	93	
. . ⊟Euteleosteomorpha	100	82	
. . . ⊟Acanthomorphata	85	72	
. . . . ⊟Euacanthomorphacea	84	71	
. . . . . ⊟Percomorphaceae	83	70	
. . . . . ⊟Ovalentaria	32	29	
. . . . . ⊟Atherinomorphae	18	15	
. . . . . ⊟Cyprinodontiformes	14	12	
. . . . . ⊟Aplocheiloidei	5	3	
. . . . . Nothobranchius furzeri	2	1	Nothobranchius furzeri hits
. . . . . ⊟Rivulidae	3	2	
. . . . . Kryptolebias marmoratus	2	1	Kryptolebias marmoratus hits
. . . . . Austrofundulus limnaeus	1	1	Austrofundulus limnaeus hits
. . . . . ⊟Cyprinodontoidei	9	9	
. . . . . ⊟Poeciliinae	7	7	
. . . . . Gambusia affinis	1	1	Gambusia affinis hits
. . . . . ⊟Poecilia	4	4	
. . . . . Poecilia latipinna	1	1	Poecilia latipinna hits
. . . . . Poecilia formosa	1	1	Poecilia formosa hits

TONS OF MATERIALS TO LEARN FROM 👻

Learn

NCBI creates a variety of educational products including courses, workshops, webinars, training materials and documentation. NCBI educational events are free and open to everyone. All NCBI educational materials are available for anyone to re-use and distribute.

Webinars & Courses

In-person courses, live webinars and webinar recordings

Conferences & Presentations

Booth exhibits and workshops at scientific conferences

Tutorials

Tutorials: Training materials in HTML, PDF and video formats

Documentation

Online manuals, handbooks, fact sheets and FAQs

Follow Us

NCBI News & Blog

New human genome annotation release with MANE Select and other improvements!

03 Jul 2019

There's a new RefSeq annotation available for the human genome, and it's quite an update! About the release Annotation release 109.20190607 is the first release of our new

Microbial Virulence in the Cloud hackathon August 13 – 15 2019

02 Jul 2019

From August 13 – 15 2019, the NCBI will run a bioinformatics hackathon on the NIH campus! We're specifically looking for folks who have experience in working with computational

GenBank release 232

01 Jul 2019

GenBank release 232.0 (6/20/2019) is now

BIOINFORMATICS CREED

Remember about biology

Do not trust the data

Use comparative approach

Use statistics

Know the limits

Remember about biology!!!