Einführung in die Angewandte Bioinformatik: Nukleotidsequenz-Datenbanken 06.05.2010

Prof. Dr. Sven Rahmann

Datenbanken am NCBI über Entrez

http://www.ncbi.nlm.nih.gov/Entrez

- NIH = National Institute of Health
- NLM = National Library of Medicine
- NCBI = National Center for Biotechnology Information

gemeinsamer Zugang über Entrez (einheitliche Eingabe von Suchanfragen)

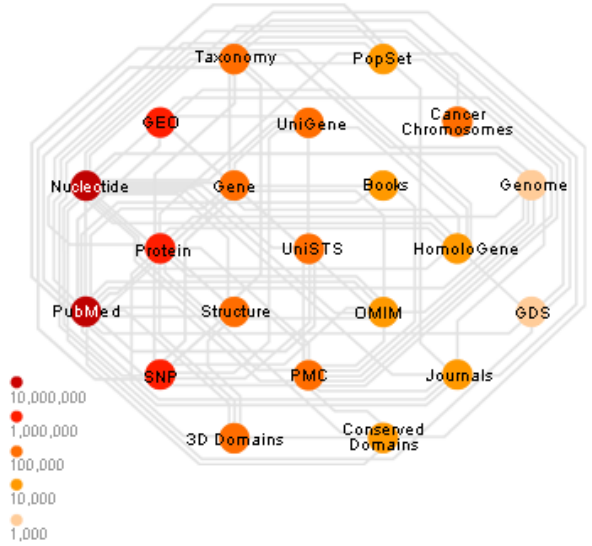
Datenbanken am NCBI über Entrez
http://www.ncbi.nlm.nih.gov/Entrez

Datenbanken-Netzwerk in Entrez
http://www.ncbi.nlm.nih.gov/Database/

unten: Nukleotid-Datenbanken

Für uns zunächst am wichtigsten:
- NCBI GenBank (NCBI Nucleotide)
- NCBI RefSeq
- NCBI Gene
- NCBI Genome

Bemerkung:
„Nucleotide“ ist eine Obermenge von GenBank, enthält z.B. auch RefSeq u.a.



The complete list of Entrez databases can be viewed in the search pull down menu.

Nucleotide Databases

dbEST
dbGSS
dbSNP
dbSTS
Nucleotide
GenBank
HomoloGene
MGC
PopSet
Probe
RefSeq
TPA
Trace Archive
UniGene
UniSTS

Warum viele Datenbanken, nicht eine (große)?

Historische Gründe; Spezialisierung; Expertenwissen

Verschiedene Personen an verschiedenen Instituten interessier(t)en sich
für verschiedene Bereiche und sammel(te)n nur dort Daten.
(Den besten Käse gibt's im Käsefachgeschäft -- nicht im Supermarkt.)

Ausfallsicherheit

Information ist physikalisch an verschiedenen Orten gespeichert.

Nur geringer Verlust an Bequemlichkeit

durch moderne WWW-basierte Systeme spielt es kaum eine Rolle,
ob im Browser angezeigte Informationen aus einer oder mehreren DBen kommen.

Achtung! Sicherheitsaspekte

Keine der öffentlichen Datenbanken unterstützt derzeit Verschlüsselung.
Jeder kann „mitlesen", für welche Sequenzen man sich interessiert.
Ungünstig für Firmen, die patent-relevante Informationen suchen.

Primärdatenbanken und Sekundärdatenbanken

Primärdatenbanken

enthalten „Rohdaten“, z.B. biologische Sequenzdaten,
dienen zur Ablage von aus Experimenten gewonnenen Daten,
können Annotationen enthalten (z.B. Herkunft, Literaturverweis)
z.B. GenBank am NCBI

Sekundärdatenbanken

enthalten aus Primärdatenbanken gewonnenes „Wissen“,
d.h. die Daten in einer Primärdatenbank werden gefiltert, überprüft, annotiert, ...
z.B. RefSeq am NCBI

Beispiel

Die Firma Celera Genomics hat um 2000 das Humangenom sequenziert,
dabei pro Tag 175000 DNA-Sequenzen (reads) à 500bp erzeugt.
Die reads speichert man in einer Primärdatenbank.
Das daraus rekonstruierte Humangenom in einer Sekundärdatenbank.

Datenbanken für Nukleotidsequenzen - Primärdatenbanken

International Nucleotide Sequence Database Collaboration
(INSDC, http://www.insdc.org)

- NCBI Nucleotide (GenBank)
  http://www.ncbi.nlm.nih.gov/Genbank/
- EMBL/EBI Nucleotide Sequence Database
  http://www.ebi.ac.uk/embl/
- DDBJ (DNA DataBase of Japan)
  http://www.ddbj.nig.ac.jp/

Gleicher Inhalt, automatischer Abgleich

Zugriff über „Metasuchmaschinen“
Entrez (NCBI), **SRS** (sequence retrieval system; EMBL),
ARSA (all-round retrieval of sequence and annotation; DDBJ)

Datenbanken für Nukleotidsequenzen - Sekundärdatenbanken

NCBI RefSeq (Referenz-Sequenz)

- genomische DNA, RNA-Transkripte
  (es gibt auch RefSeq für Proteine, siehe später)

- stabile Basis für weitere Annotationen
  - wo liegen Gene, Transkripte, ... im Genom,
  - wo gibt es Mutationen,
  - wo konservierte und variable Segmente verglichen mit verwandten Organismen?

- breite taxonomische Abdeckung: es gibt RefSeqs für
  - Eurkaryoten, Prokaryoten, Viren

- seltene Änderungen

Datenbanken für Nukleotidsequenzen - Sekundärdatenbanken

NCBI Gene (Informationen zu Genen)

- Verweise auf zugehörige Nukleotid-Sequenzen
- Funktion eines Gens
- verschiedene Isoformen eines Gens
- Proteine dazu

NCBI Genome (komplette Genome)

- Liste aller öffentlich bisher sequenzierten Genome, viele Bakterien, verschiedene Strains verfügbar

Wichtige Unterschiede
Primärdatenbank GenBank vs. Sekundärdatenbank RefSeq

GenBank	**RefSeq**
Not curated	Curated
Author submits	NCBI creates from existing data
Only author can revise	NCBI revises as new data emerge
Multiple records for same loci common	Single records for each molecule of major organisms
Records can contradict each other	
No limit to species included	Limited to model organisms
Data exchanged among INSDC members	Exclusive NCBI database
Akin to primary literature	Akin to review articles
Proteins identified and linked	Proteins and transcripts identified and linked
Access via NCBI Nucleotide databases	Access via Nucleotide & Protein databases

Quelle: http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=handbook.section.GenBank_ASM

NCBI Entrez GenBank / Nucleotide

Primärdatenbank,
nicht kuriert, daher viele doppelte Einträge, „Müll“

Publikation neuer Nukleotidsequenzen in GenBank
bevor man in einer relevanten molekularbiologischen Zeitschrift veröffentlicht.

Exponentielles Wachstum seit den 1990er Jahren.
Feb. 2008: 85,759,586,764 nt in 82,853,685 Sequenz-Einträgen
Apr. 2009: 102,980,268,709 nt in 103,335,421 Sequenz-Einträgen
Apr. 2010: 114,348,888,771 nt in 119,112,251 Sequenz-Einträgen

Quelle: Release Notes der aktuellen GenBank-Version unter
ftp://ftp.ncbi.nih.gov/genbank/gbrel.txt

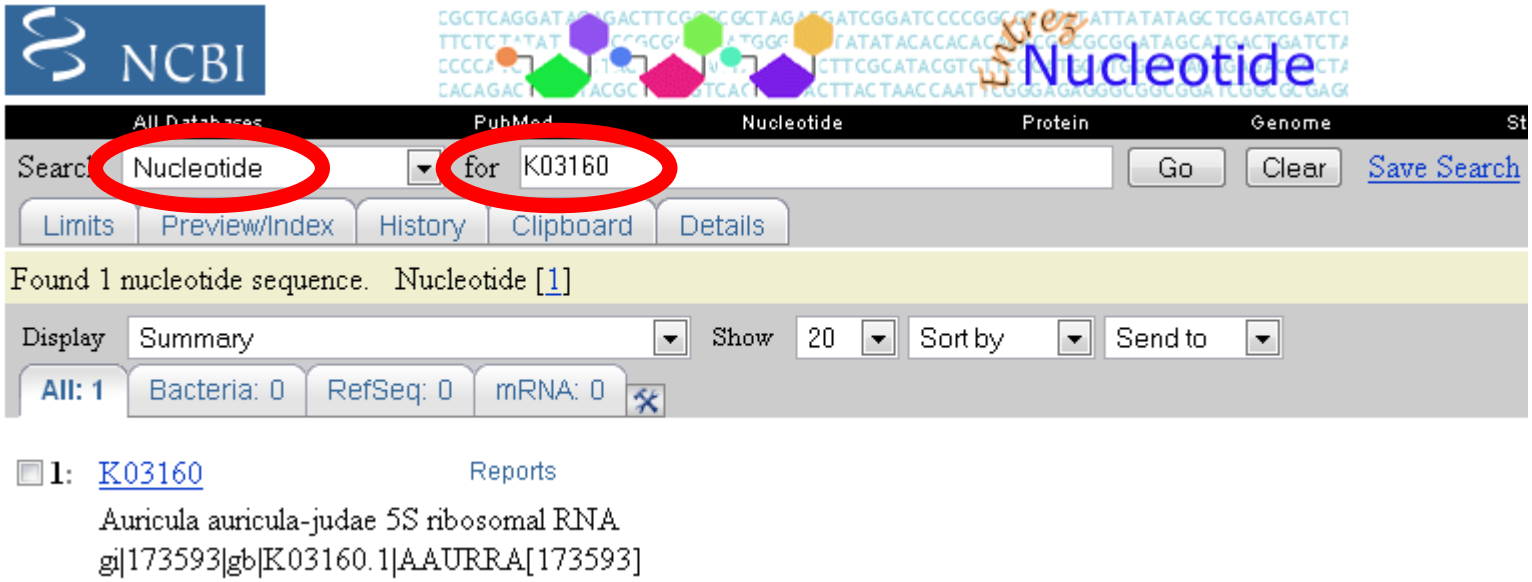
Zugriffsnummern (Accession Numbers; ACs)

Nummer, die einen Datenbankeintrag eindeutig identifiziert („Schlüssel“)
Beispiel: PubMed ID; Feld [PMID] bei der PubMed-Datenbanksuche.
Kennt man die AC, braucht man keine komplexen Suchanfragen.

ACs werden benutzt, um Datenbanken zu verbinden.
Beispiel: Bei einer (codierenden) DNA-Sequenz in einer Nukleotiddatenbank findet man die AC der zugehörigen Proteinsequenz(en) für eine Proteindatenbank.

Verwendung:
- Manuell – notiere ACs auf einem Zettel, suche damit in anderen Datenbanken
- Per Link – direkter WWW-Link zu den referenzierten Datenbanken
- Datenintegrationssysteme – fragen automatisch mehrere Datenbanken ab, stellen ggf. Ergebnisse übersichtlich zusammen

Suche in NCBI GenBank mit Accession Number



Eingabe einer Accession Number führt sofort zum gewünschten Eintrag.
In diesem Fall (da „K03160“ niemals sonst als Stichwort auftritt)
muss man nicht explizit angeben, dass es sich um [AC] handelt.

Ergebnis im GenBank-Format

1: K03160. Reports Auricula auricula...[gi:173593]

Features Sequence

```
LOCUS       AAURRA                     118 bp    rRNA    linear   PLN 13-DEC-1995
DEFINITION  [illegible] auricula-judae 5S ribosomal RNA.
ACCESSION   K03160
VERSION     [illegible] GI:173593
KEYWORDS    5S ribosomal RNA; ribosomal RNA.
SOURCE      Auricularia auricula-judae (ear fungus)
  ORGANISM  Auricularia auricula-judae
            Eukaryota; Fungi; Dikarya; Basidiomycota; Agaricomycotina;
            Agaricomycetes; Auriculariales; Auriculariaceae; Auricularia.
REFERENCE   1  (bases 1 to 118)
  AUTHORS   Huysmans,E., Dams,E., Vandenberghe,A. and De Wachter,R.
  TITLE     The nucleotide sequences of the 5S rRNAs of four mushrooms and
            their use in studying the phylogenetic position of basidiomycetes
            among the eukaryotes
  JOURNAL   Nucleic Acids Res. 11 (9), 2871-2880 (1983)
   PUBMED   6856478
COMMENT     Original source text: Auricularia auricula-judae rRNA.
FEATURES             Location/Qualifiers
     source          1..118
                     /organism="Auricularia auricula-judae"
                     /mol_type="rRNA"
                     /db_xref="taxon:29892"
     rRNA            1..>118
                     /product="5S ribosomal RNA"
ORIGIN
        1 atccacggcc ataggactct gaaagcactg catcccgtcc gatctgcaaa gttaaccaga
       61 gtaccgccca gttagtacca cggtgggggа ccacgcggga atcctgggtg ctgtggtt
//
```

Ein GenBank-Eintrag
- beginnt mit LOCUS
- endet mit //
- Sequenz hinter ORIGIN
- Annotation: FEATURES

GeneBank-Format (flat file) – kann man herunterladen

```
LOCUS       AAURRA        118 bp ss-rRNA            RNA       16-JUN-1986
DEFINITION  A.auricula-judae (mushroom) 5S ribosomal RNA.
ACCESSION   K03160
VERSION     K03160.1  GI:173593
KEYWORDS    5S ribosomal RNA; ribosomal RNA.
SOURCE      A.auricula-judae (mushroom) ribosomal RNA.
  ORGANISM  Auricularia auricula-judae
            Eukaryota; Fungi; Eumycota; Basidiomycotina; Phragmobasidiomycetes;
            Heterobasidiomycetidae; Auriculariales; Auriculariaceae.
REFERENCE   1  (bases 1 to 118)
  AUTHORS   Huysmans,E., Dams,E., Vandenberghe,A. and De Wachter,R.
  TITLE     The nucleotide sequences of the 5S rRNAs of four mushrooms and
            their use in studying the phylogenetic position of basidiomycetes
            among the eukaryotes
  JOURNAL   Nucleic Acids Res. 11, 2871-2880 (1983)
FEATURES             Location/Qualifiers
     rRNA            1..118
                     /note="5S ribosomal RNA"
BASE COUNT       27 a     34 c     34 g     23 t
ORIGIN      5' end of mature rRNA.
        1 atccacggcc ataggactct gaaagcactg catcccgtcc gatctgcaaa gttaaccaga
       61 gtaccgccca gttagtacca cggtggggga ccacgcggga atcctgggtg ctgtggtt
//
LOCUS       ABCRRAA       118 bp ss-rRNA            RNA       15-SEP-1990
...
```

Suchmöglichkeiten in GenBank (und anderen NCBI-DBs)

[ACCN] – sucht nach Accession Number
[SLEN] – sucht nach bestimmten Sequenzlängen
[ORGN] – sucht nach Sequenzen im angegebenen Organismus

Bereichssuche (:)

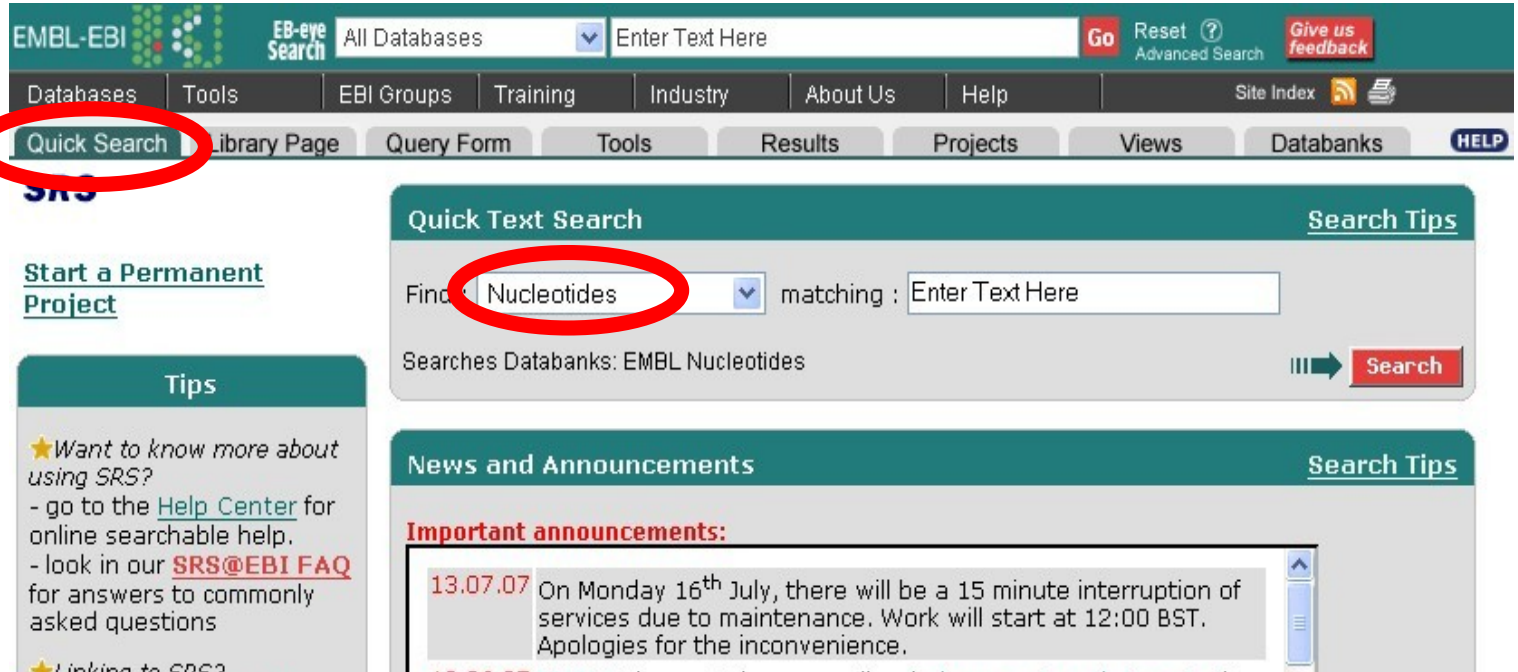
Bei numerischen Eingaben kann mit einem Doppelpunkt (:)
ein ganzer Bereich durchsucht werden.
Beispiel: 200:222 [SLEN]
findet Sequenzen der Länge 200 bis 222

Wildcard-Suche (*)

Ist das Wortende nicht genau bekannt oder soll offen gelassen werden,
kann es mit einem Stern (*) abgeschnitten werden.
Beispiel: HUM* [ORGN]
findet Sequenzen in *Homo sapiens* (human) oder *Humulus lupulus* (Hopfen)

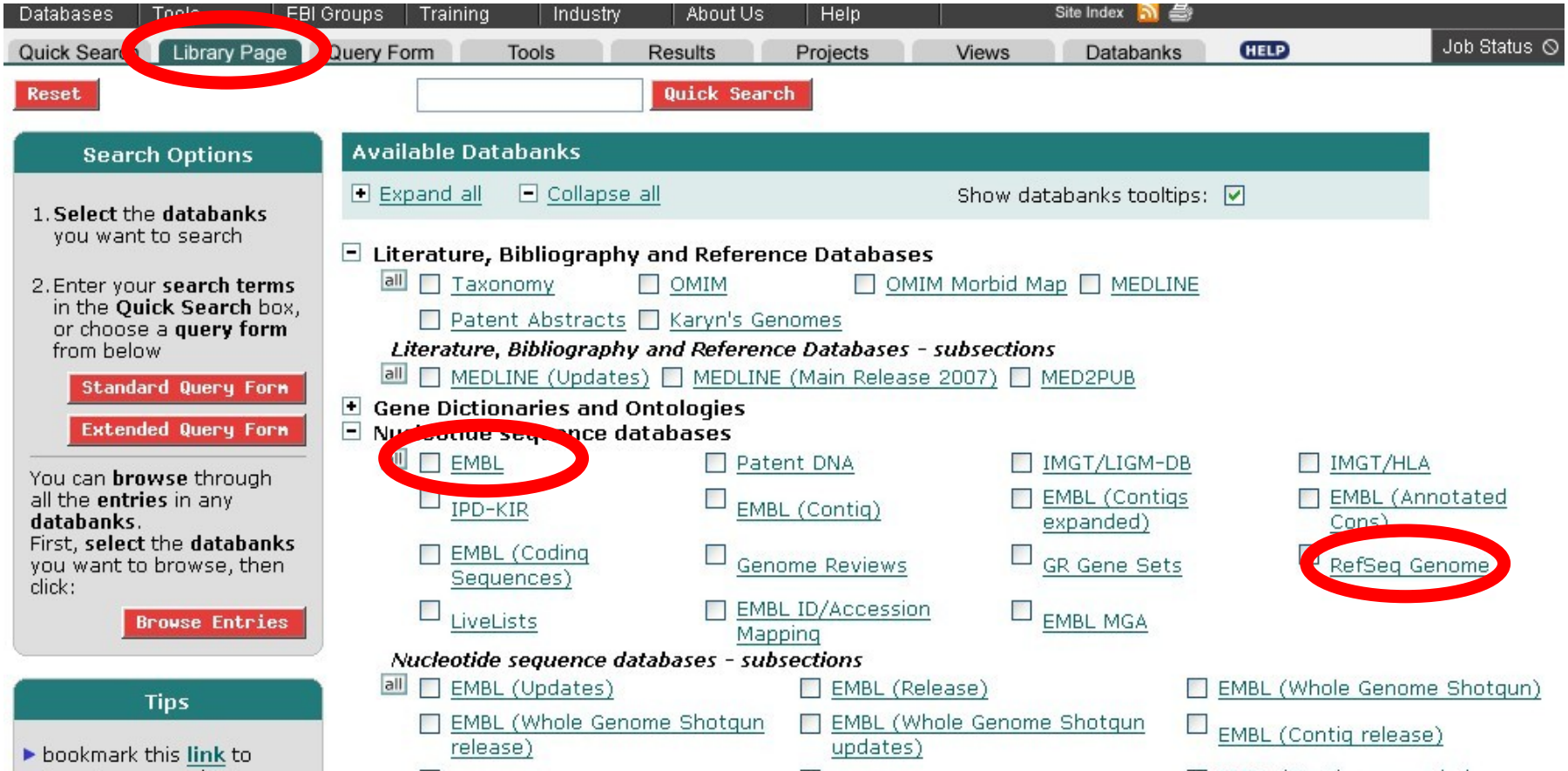
Suchmöglichkeiten in EMBL Nucleotide mit SRS

EMBL Nucleotide ist inhaltlich eine Kopie von GenBank.
Statt mit Entrez durchsucht man sie mit SRS (Sequence Retrieval System)
URL: http://srs.ebi.ac.uk/
- schnelle Sitchwortsuche über „Quick Text search“
- genaue Eingrenzung durch Klick auf Reiter „Library page“, „Query Form“



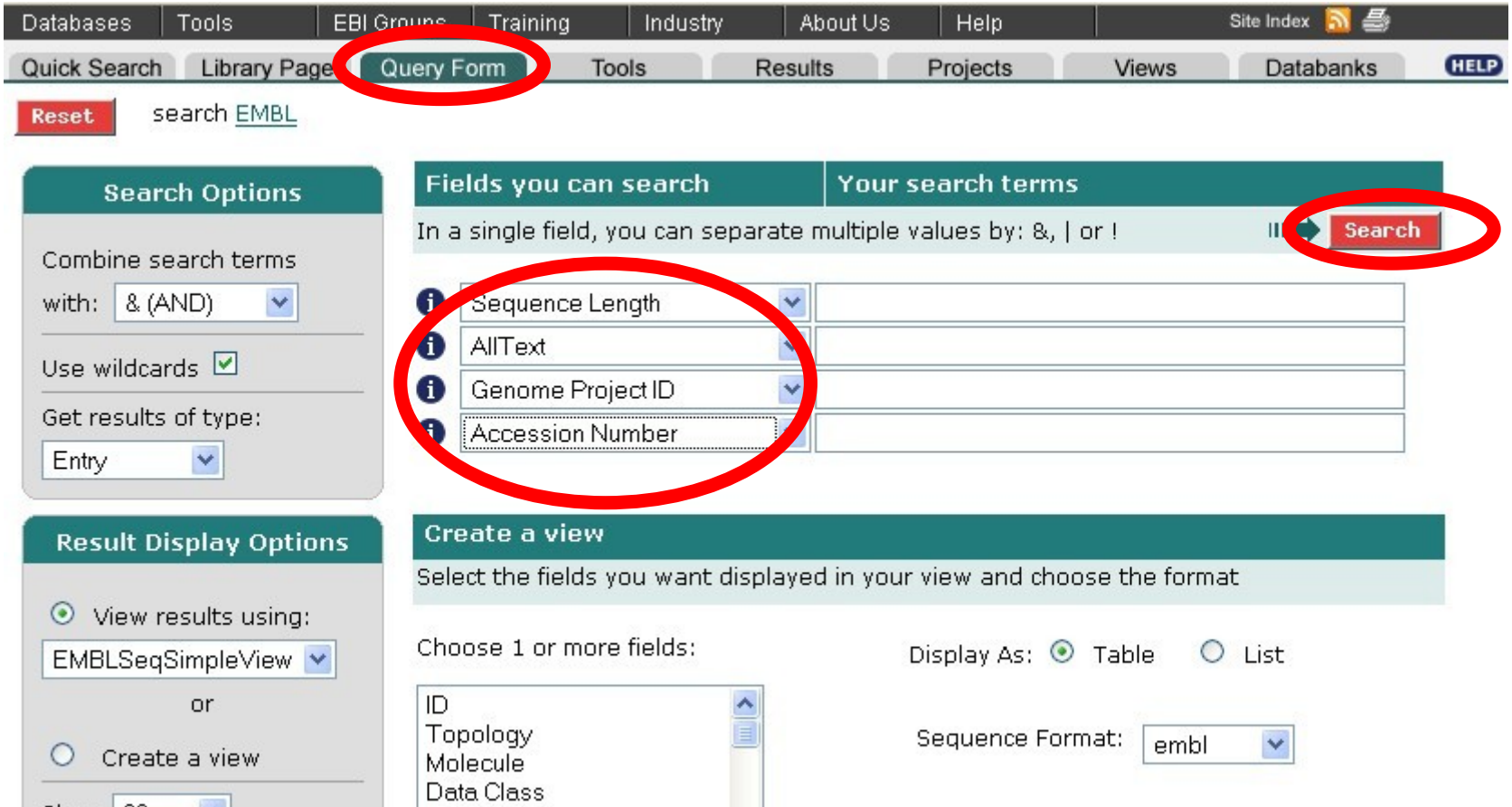
EMBL Nucleotide mit SRS – Library Page

Auswahl der zu durchsuchenden Datenbanken



EMBL Nucleotide mit SRS – Query Form

Einstellung der Suchparameter; dann Klick auf „Search“



EMBL Nucleotide mit SRS – Ergebnisseite

General Information	
Primary Accession #	K03160
Accession #	K03160
SRS Entry ID	EMBL:K03160 (*formerly EMBL:AARRA*)
Molecule Type	linear rRNA
Sequence Length	118
Entry Division	FUN (*Fungi*)
Entry Data Class	STD (*Standard*)
Sequence Version	K03160.1
Creation Date	17-JAN-1991
Modification Date	04-MAR-2000
EMBL-SVA	K03160

Description	
Description	Auricula auricula-judae 5S ribosomal RNA.
Keywords	5S ribosomal RNA; ribosomal RNA.;
Organism	Auricularia auricula-judae (ear fungus)
Organism Classification	Eukaryota; Fungi; Dikarya; Basidiomycota; Agaricomycotina; Agaricomycetes; Auriculariales; Auriculariaceae; Auricularia.

References

1. Huysmans,E.; Dams,E.; Vandenberghe,A.; De Wachter,R.;
The nucleotide sequences of the 5S rRNAs of four mushrooms and their use in studying the phylogenetic position of
Nucleic Acids Res. 11(9):2871-2880 (1983)

DOI	10.1093/nar/11.9.2871
PubMed	6856478 CiteXplore
Position	1-118

Database Cross-references

RFAM RF00001.

Features

Key	Location	Qualifier	Value
source	1..118	organism	Auricularia auricula-judae
		mol_type	rRNA
		db_xref	taxon:29892
rrna	1..>118	product	5S ribosomal RNA

Sequence

Characteristics	**Length:** 118 BP, **A Count:**27, **C Count:**34, **G Count:**34, **T Count:**23, **Others Count:**0
Sequence	>embl|K03160|K03160 Auricula auricula-judae 5S ribosomal RNA.

Ergebnis der Suche nach Accession Number K03160:

Inhalt wie bei NCBI-Entrez, nur andere Darstellung

EMBL-Format (flat file) mit SRS

2-Buchstaben-Kürzel identifizieren die Elemente des Eintrags

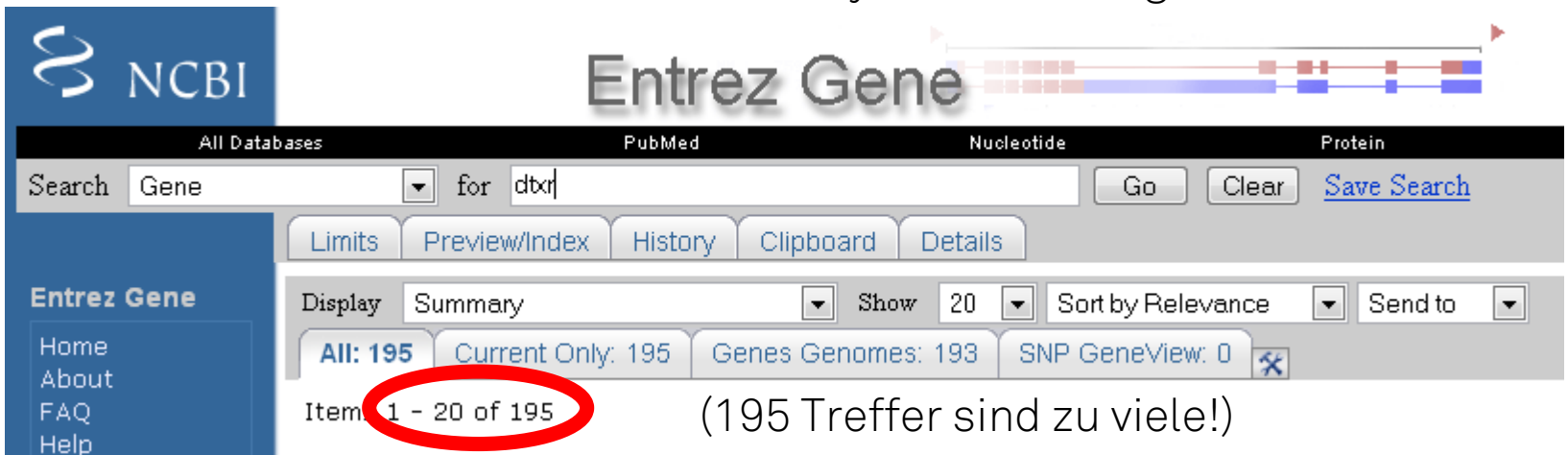
```
ID   K03160; SV 1; linear; rRNA; STD; FUN; 118 BP.
AC   K03160;
DT   17-JAN-1991 (Rel. 26, Created)
DT   04-MAR-2000 (Rel. 63, Last updated, Version 7)
DE   Auricula auricula-judae 5S ribosomal RNA.
KW   5S ribosomal RNA; ribosomal RNA.
OS   Auricularia auricula-judae (ear fungus)
OC   Eukaryota; Fungi; Dikarya; Basidiomycota; Agaricomycotina; Agaricomycetes;
OC   Auriculariales; Auriculariaceae; Auricularia.
RN   [1]
RP   1-118
RX   DOI; 10.1093/nar/11.9.2871
RX   PUBMED; 6856478.
RA   Huysmans E., Dams E., Vandenberghe A., De Wachter R.;
RT   "The nucleotide sequences of the 5S rRNAs of four mushrooms and their use
RT   in studying the phylogenetic position of basidiomycetes among the
RT   eukaryotes";
RL   Nucleic Acids Res. 11(9):2871-2880(1983).
XX
DR   RFAM; RF00001.
FH   Key             Location/Qualifiers
FH
FT   source          1..118
FT                   /organism="Auricularia auricula-judae"
FT                   /mol_type="rRNA"
FT                   /db_xref="taxon:29892"
FT   rRNA            1..>118
FT                   /product="5S ribosomal RNA"
XX
SQ   Sequence 118 BP; 27 A; 34 C; 34 G; 23 T; 0 other;
     atccacggcc ataggactct gaaagcactg catcccgtcc gatctgcaaa gttaaccaga        60
     gtaccgccca gttagtacca cggtggggga ccacgcggga atcctgggtg ctgtggtt         118
//
```

NCBI Entrez Gene (Sekundärdatenbank)

Motivation für Entrez Gene

GenBank enthält alle möglichen Arten von Nukleotid-Sequenzen.
Nicht jedes „Stück DNA“ im Genom gehört zu einem Gen.
Ein Gen kann durch viele Einträge in GenBank repräsentiert sein.
Man benötigt eine Gen-zentrierte „Sicht“ auf diese Sequenzen;
Gene sind ein viel häufigerer Startpunkt für eine Suche als einzelne Sequenzen.

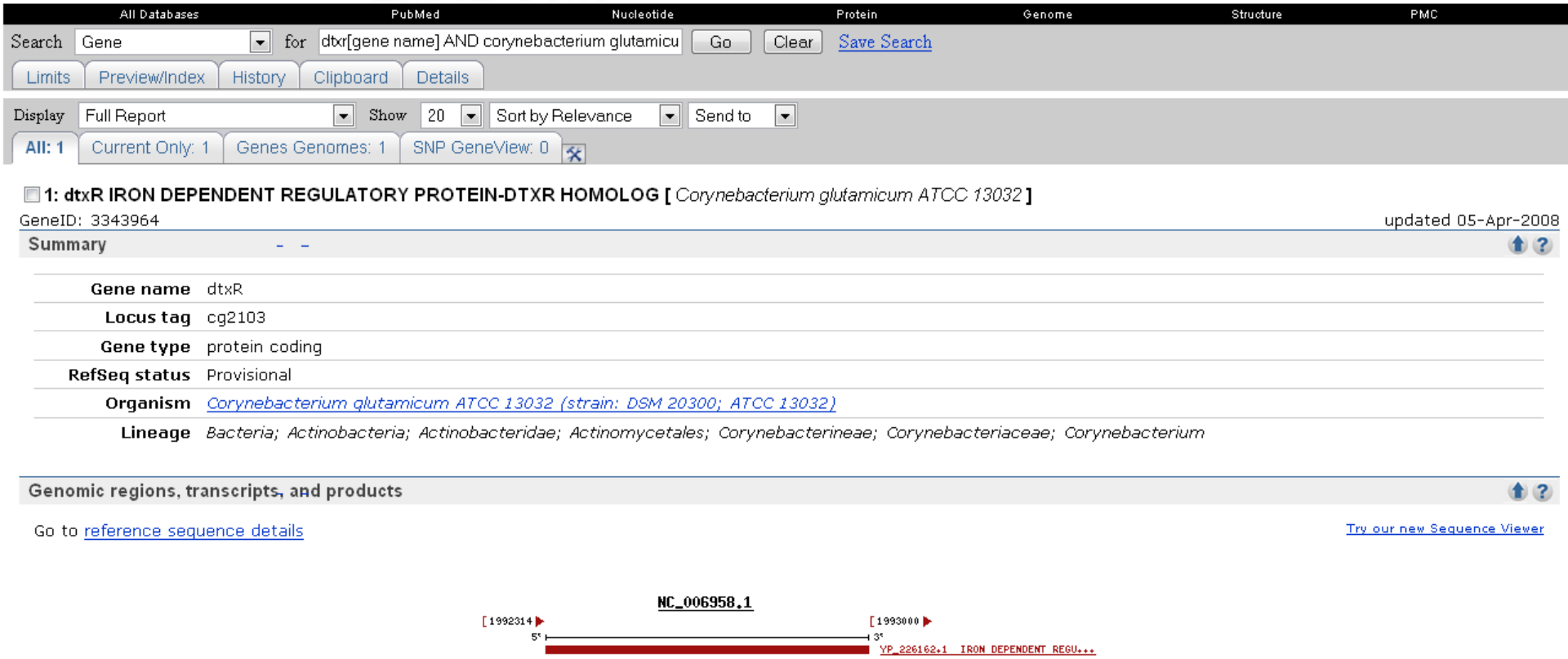
Beispiel:
In einem Artikel wird das *dtxR*-Gen in *Corynebacterium glutamicum* erwähnt.



(195 Treffer sind zu viele!)

Suche mit Feldnamen in Entrez Gene

[gene name] – Name des Gens
[organism] = [orgn] – Organismus, in dem man nach Genen sucht
Beispiel: dtxr[gene name] AND Corynebacterium glutamicum[organism]



Suche mit Feldnamen – Hinweise

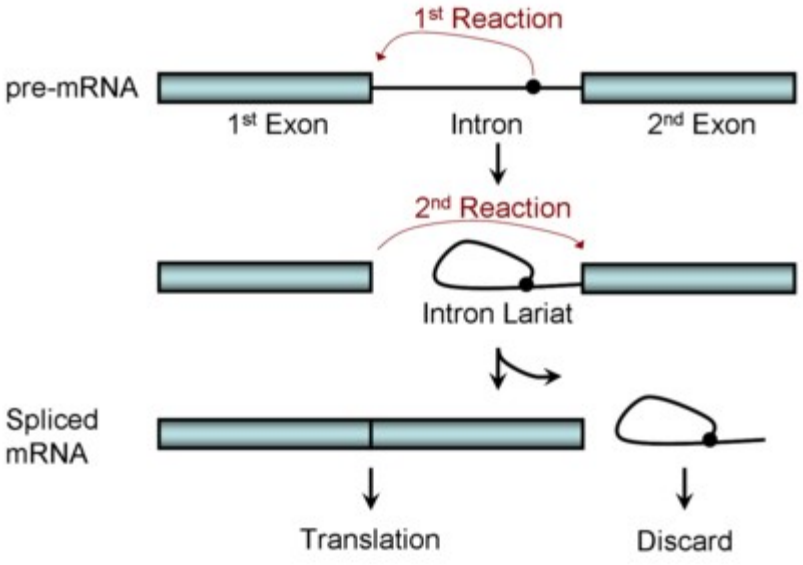
Feldnamen sind hilfreich zur Präzisierung der Suche.
Liste aller Feldnamen: Hilfe-Seiten am NCBI
Beispiele auch in den Übungsaufgaben.

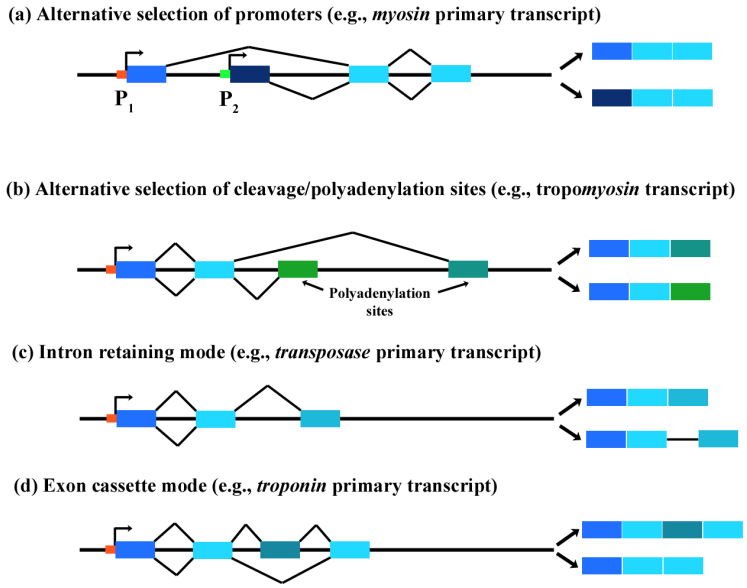
Erinnerung: Boole'sche Operatoren AND, OR, NOT
Entrez wertet ungewöhnlich aus: von links nach rechts
Daher Klammern setzen, um Gruppen von Bedingungen abzugrenzen.

Beispiel (in der NCBI Gene-Datenbank):
Gene der Maus auf Chromosom 1 oder 2, die mit Krebs in Verbindung stehen?
Falsch: cancer [dis] AND mouse [orgn] AND 1 [chr] OR 2 [chr] – 128588 Ergebnisse
Richtig: cancer [dis] AND mouse [orgn] AND (1 [chr] OR 2 [chr]) – 7 Ergebnisse
Richtig: 1 [chr] OR 2 [chr] AND cancer [dis] AND mouse [orgn] – 7 Ergebnisse
Richtig: (1 [chr] OR 2 [chr]) AND cancer [dis] AND mouse [orgn] – 7 Ergebnisse

Isoformen von Genen (alternatives Splicing)

Transkription von Protein-codierenden Genen:
Genomische DNA wird durch RNA-Polymerase abgelesen und in mRNA kopiert.
Spleißen (splicing): Introns werden aus mRNA entfernt.
Durch alternatives Splicing können aus einem Transkript
verschiedene mRNAs (Isoformen desselben Gens) und damit Proteine entstehen.





Bildquellen: Wikipedia

Alternatives Splicing in NCBI Entrez Gene

Ergebnis-Seiten zeigen alle bekannten Isoformen eines Gens.

Beispiel: Homo sapiens, Myosin heavy chain 14

1: MYH14 myosin, heavy chain 14 [*Homo sapiens* **]**

GeneID: 79784 updated 25-Apr-2008

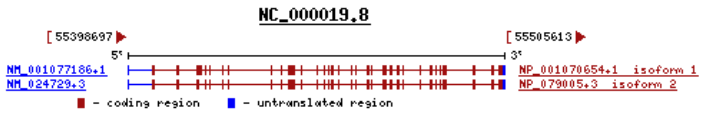
Summary

Official Symbol	MYH14 provided by HGNC
Official Full Name	myosin, heavy chain 14 provided by HGNC
Primary source	HGNC:23212
See related	Ensembl:ENSG00000105357; HPRD:10543; MIM:608568
Gene type	protein coding
RefSeq status	Reviewed
Organism	*Homo sapiens*
Lineage	*Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo*
Also known as	DFNA4; MHC16; myosin; FP17425; FLJ13881; FLJ43092; KIAA2034; NMHC-II-C; DKFZp667A1311
Summary	This gene encodes a member of the myosin superfamily. Myosins are actin-dependent motor proteins with diverse functions including regulation of cytokinesis, cell motility, and cell polarity. Mutations in this gene result in one form of autosomal dominant hearing impairment. Multiple transcript variants encoding different isoforms have been found for this gene.

Genomic regions, transcripts, and products

Go to reference sequence details

Try our new Sequence Viewer

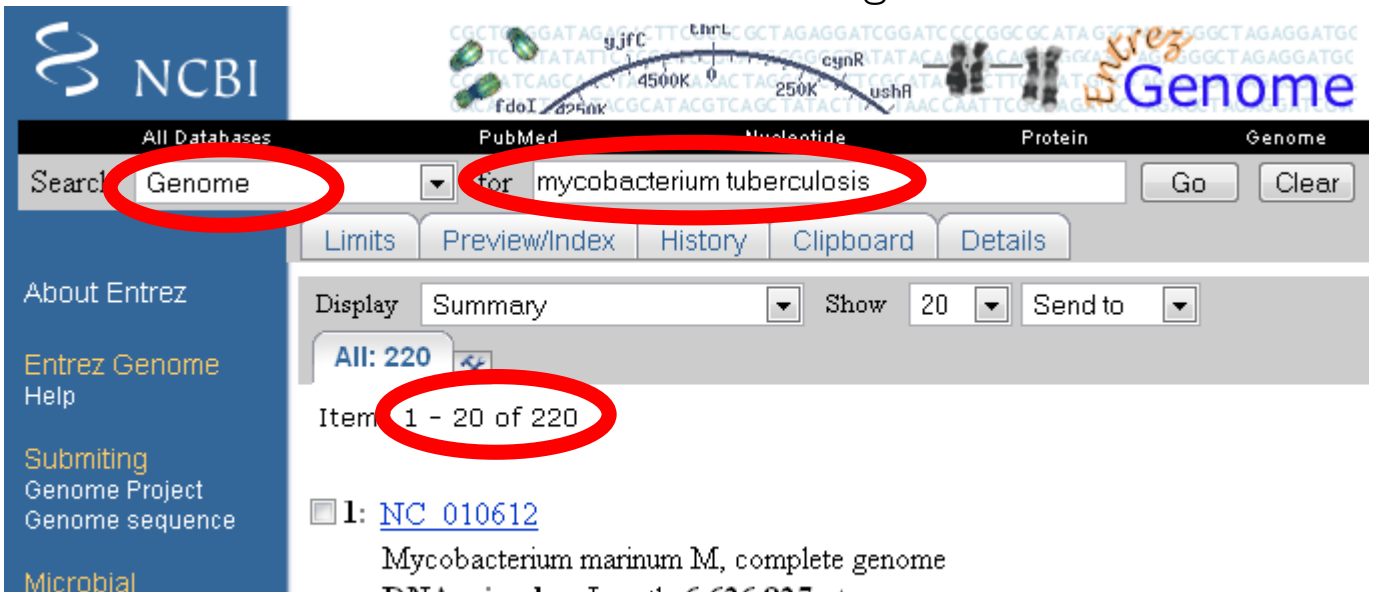


NCBI Entrez Genome (Sekundärdatenbank)

Motivation für Entrez Genome

Es gibt mittlerweile viele komplett sequenzierte Genome;
d.h. die gesamte Erbinformation vieler Spezies ist mittlerweile bekannt.
Bei Bakterien sind sogar oft mehrere Stämme (strains) sequenziert.

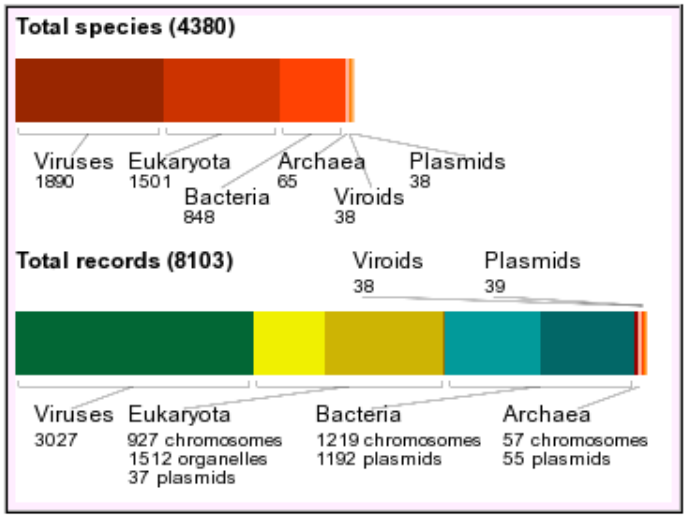
Beispiel:
Wie viele Stämme des Tuberkulose-Erregers sind bereits bekannt?



220 ist falsch!
[orgn] verwenden!
Lösung: 21

NCBI Entrez Genome – Liste aller Genome

Alle Genome in 'NCBI Genome' lassen sich auflisten:
unter http://www.ncbi.nlm.nih.gov/sites/entrez?db=genome



Inhalt der Genome – Datenbank
(Quelle: NCBI; Mai '08)

Durch Klick auf 'Bacteria' kommt man zu einer Liste, in der man auch)
alle sequenzierten (öffentlich zugänglichen) Stämme von *M. tuberculosis* findet:

Mycobacterium tuberculosis CDC1551	chromosome	NC_002755	4403837 nt
Mycobacterium tuberculosis F11	chromosome	NC_009565	4424435 nt
Mycobacterium tuberculosis H37Ra	chromosome	NC_009525	4419977 nt
Mycobacterium tuberculosis H37Rv	chromosome	NC_000962	4411532 nt

Zusammenfassung

Nukleotid-Sequenzdatenbanken

- Primär: NCBI Entrez Nucleotide (GenBank) = EMBL Nucleotide = DDBJ
- Sekundär: Entrez RefSeq, Entrez Gene, Entrez Genome

Suche in Sequenzdatenbanken

- verschiedene Systeme (Entrez, SRS), gleiche Möglichkeiten
- Feldnamen verwenden zur Eingrenzung
- Bereichs-Operator „:“ und Wildcard-Operator „*“
- Boole'sche Operatoren AND, OR, NOT: Klammern setzen!

Fortsetzung

Statistik mit R

Zugriff auf Elemente von Vektoren

Vektoren werden mit eckigen Klammern [] indiziert.
Indizierung beginnt bei 1, nicht bei 0 !
Man kann auch mit Vektoren indizieren!

```
x = seq(0, 10, by=0.5)      # erzeugt 0, 0.5, 1, ..., 9.5, 10
x[0]                        # gibt es nicht
x[1]                        # 0
x[length(x)]                # 10
x[seq(1,length(x),3)]       # jedes dritte Element von x
```

Tests von Vektoren, logische Indizierung

Elemente von Vektoren können auf Eigenschaften getestet werden.
Wie viele von 20 gleichverteilten Zufallszahlen auf [0,1] sind >0.5 ?

```
x = runif(20)          # 20 Zufallszahlen zwischen 0 und 1
gr = x>0.5             # Vektor aus TRUE und FALSE
sum(gr)                # Anzahl der TRUE-Werte
y = x[gr]              # x-Werte >0.5 nach y extrahieren
mean(y)                # Durchschnitt dieser Werte
mean(x[x>0.5])         # dasselbe! Lies:
                       # Mittelwert der x-Werte, für die x>0.5
```

Merke:
Vektor-Indizierung kann auf zwei Arten erfolgen:

- numerisch (mittels Zahlenvektor: x[c(1,2,3)])
- logisch (mittels T/F-Vektor: x[c(T,T,T,F,F,F)])

Runde und Eckige Klammern

- Runde Klammern: Funktionsaufruf
  - mean ist eine Funktion; wird angewendet auf einen Vektor
  - mean(x) liefert Mittelwert des Vektors x
  - runif ist eine Funktion; wird angewendet auf eine natürliche Zahl n
  - runif(n) liefert n zufällige Zahlen im Intervall [0,1[.

- Eckige Klammern: Indizierung
  - z.B. sei x ein Vektor; das k-te Element bekommt man mit x[k]
  - Wir wissen: Auch der Index kann ein Vektor sein;
  - dieser kann durch eine Funktion wie seq oder c erzeugt werden:
  - x[seq(1,10,2)] oder x[c(1,3,5,7,9)]

Funktionsaufruf mit benannten Parametern

Es ist Konvention, einer Funktion erst die nötigen Daten zu übergeben;
danach benannte Parameter, die das Verhalten der Funktion verändern.

Beispiel: Plot mit Linien und Punkten in Blau mit Titel:

```
plot(x,y, type="o", col="blue", main="Ein Funktionsplot")
```

Beispiel: Logarithmische Achsen
Achsen können mit logarithmischer Skala darzustellen.
Dies geschieht (für die y-Achse) durch Angabe des Parameters log="y".
Monome der Form $y=cx^n$ werden bei logarithmischen Achsen zu Geraden.

```
x = seq(1, 5, by=1/16)
y = 5 * x**3
plot(x,y, type="o")
plot(x,y, type="o", log="xy")  # beide Achsen logarithmisch
```