CAS STNext® COFFEE LECTURE

Searching for biological species in CAS STNext®

Jan Baur, ACS International

© 2023 American Chemical Society. All rights reserved.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

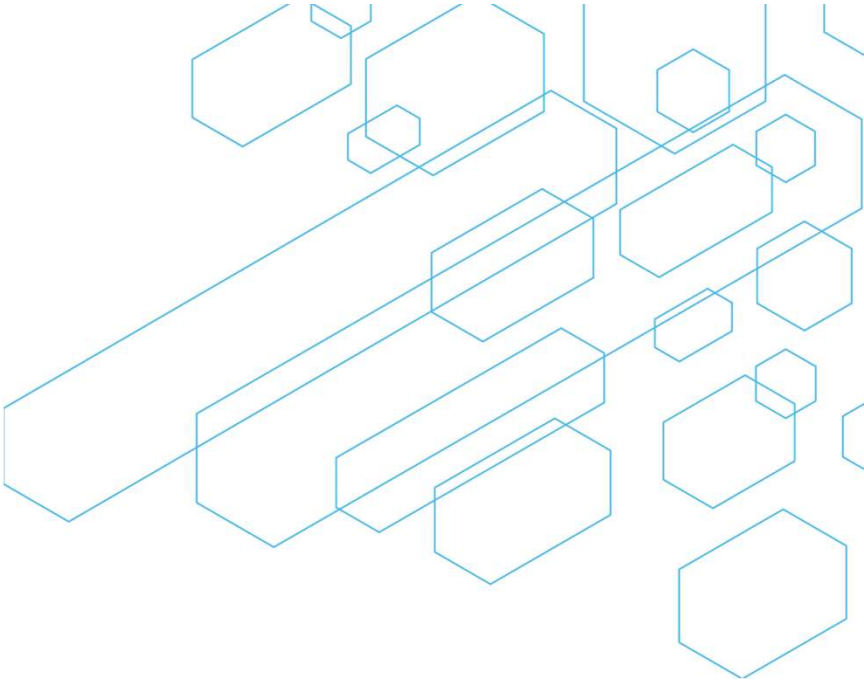
CAS
A division of the American Chemical Society

Agenda

Basics

Relevant Databases

Hierarchies and search examples



© 2023 American Chemical Society. All rights reserved.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

The same organism might be known under several names

Fermipan

Baker's Yeast

Fermirouge

Biosaf

Saccharomyces cerevisiae

S. cerevisiae

Cerevisane

Romeo

Faex medicinalis

© 2023 American Chemical Society. All rights reserved.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

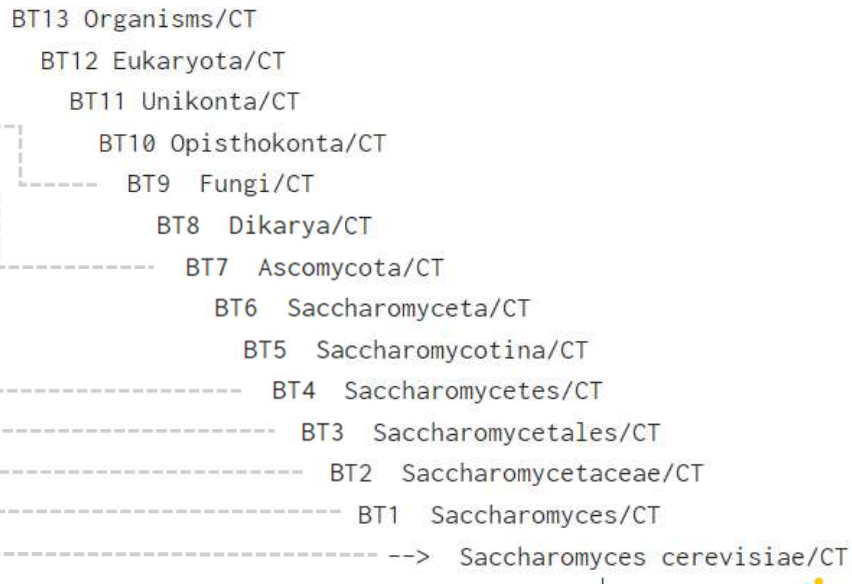
Is yeast a bacterium? It's a fungus

Every organism finds its place in the biological taxonomical tree

Taxonomic tree of *Saccharomyces cerevisiae* in the Catalogue of Life

https://www.catalogueoflife.org/

Taxonomic tree in an STN database: CTs in CAplus

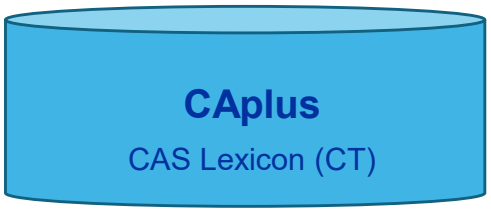


© 2023 American Chemical Society. All rights reserved.

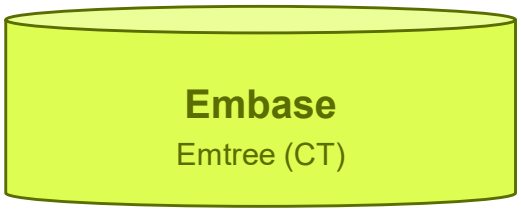
FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

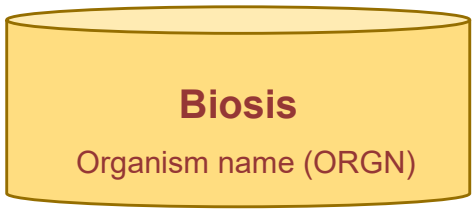
Select STNext databases with organism hierarchies



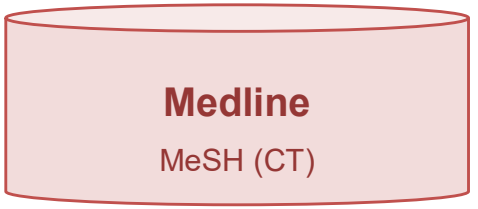
Chemistry, biochemistry and related fields



Pharmacology, biomedicine and medical devices

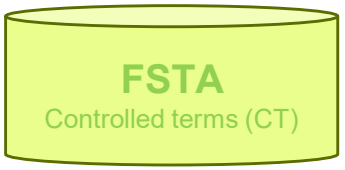


Worldwide literature on biological and biomedical topics



Worldwide literature on biomedical topics

Agriculture and related fields



Food science and technology

© 2023 American Chemical Society. All rights reserved.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Expand a taxonomical term in the CT field

Example: Plant species and model organism *Nicotiana benthamiana* in CAplus

```
=> e Nicotiana benthamiana+all/ct

E1         6228   BT9  Organisms/CT
E2        22128     BT8  Eukaryota/CT
E3        78047       BT7  Plantae/CT
E4          153         BT6  Spermatophyta/CT
E5         4183           BT5  Magnoliophyta/CT
E6         4674             BT4  Magnoliopsida/CT
E7           55               BT3  Solanales/CT
E8         3561                 BT2  Solanaceae/CT
E9         1268                   BT1  Nicotiana/CT
E10       11066                     -->  Nicotiana benthamiana/CT
. . .
```


© 2023 American Chemical Society. All rights reserved.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Enter the taxonomy at the genus level in CAplus

Check which tobacco species are discussed most

```
=> e Nicotiana+nt/ct

E15        1268    -->  Nicotiana/CT
                      NOTE This heading is used for studies of the genus
                           Nicotiana and unidentified species. Commercial tobacco
                           crop is indexed at Nicotiana tabacum, unless otherwise
                           indicated in the document.
E16           8      NT1  Nicotiana acaulis/CT
E17          23      NT1  Nicotiana acuminata/CT
E18           5        NT2  Nicotiana acuminata multiflora/CT
. . .
E28       11066      NT1  Nicotiana benthamiana/CT
. . .
E111      45353      NT1  Nicotiana tabacum/CT
E112          1        NT2  Nicotiana tabacum aurea/CT
E113          0        NT2  Nicotiana tabacum gossii/CT
E114          0        NT2  Nicotiana tabacum macrophylla/CT
E115          0        NT2  Nicotiana tabacum paraguay/CT
E116         42        NT2  Nicotiana tabacum samsun/CT
E117          1        NT2  Nicotiana tabacum trapezond/CT
E118        124        NT2  Nicotiana tabacum xanthi/CT
. . .
```

CTs in CAplus are indexed on the highest level of specificity and **there's no automatic up-posting**.

I.e. the genus Nicotiana has less records compared to the prominent model organisms *N. benthamiana* and *N. tabacum*.

Two species are highly discussed: *N. benthamiana* and *N. tabacum*

© 2023 American Chemical Society. All rights reserved.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Search thesaurus terms using relationship codes in CAplus

```
=> e Nicotiana benthamiana+all/ct

E1         6229   BT9  Organisms/CT
E2        22144     BT8  Eukaryota/CT
E3        78076       BT7  Plantae/CT
E4          154         BT6  Spermatophyta/CT
E5         4185           BT5  Magnoliophyta/CT
E6         4674             BT4  Magnoliopsida/CT
E7           55               BT3  Solanales/CT
E8         3561                 BT2  Solanaceae/CT
E9         1268                   BT1  Nicotiana/CT
E10       11073                     -->  Nicotiana benthamiana/CT
                                      HNTE Valid heading during volume 136
                                           (2002) to present.
E11                                   OLD  Tobacco (L) N. benthamiana/CT
E12                                   OLD  Tobacco (L) Nicotiana benthamiana/CT
E13          239                      OLD  Tobacco (Nicotiana benthamiana)/CT
E14                                   UF   N. benthamiana/CT. . .
```

```
L1 27 S NICOTIANA BENTHAMIANA+NT,OLD/CT AND
(ACE2 OR ACE 2 OR angiotensin converting enzyme 2)
```

Searches for the indexed species name, including previously used CTs (+OLD). No narrower terms exist.

© 2023 American Chemical Society. All rights reserved.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Search thesaurus terms using relationship codes in CAplus

```
=> e Nicotiana+nt/ct

E15        1268    -->  Nicotiana/CT
                      NOTE This heading is used for studies of the genus
                           Nicotiana and unidentified species. Commercial tobacco
                           crop is indexed at Nicotiana tabacum, unless otherwise
                           indicated in the document.
E16           8      NT1  Nicotiana acaulis/CT
E17          23      NT1  Nicotiana acuminata/CT
E18           5        NT2  Nicotiana acuminata multiflora/CT
. . .
E28       11066      NT1  Nicotiana benthamiana/CT
. . .
E111      45353      NT1  Nicotiana tabacum/CT
E112          1        NT2  Nicotiana tabacum aurea/CT
E113          0        NT2  Nicotiana tabacum gossii/CT
E114          0        NT2  Nicotiana tabacum macrophylla/CT
E115          0        NT2  Nicotiana tabacum paraguay/CT
E116         42        NT2  Nicotiana tabacum samsun/CT
E117          1        NT2  Nicotiana tabacum trapezond/CT
E118        124        NT2  Nicotiana tabacum xanthi/CT
. . .
```

```
L2  44  S NICOTIANA+NT,OLD/CT AND (ACE2 OR
ACE 2 OR angiotensin converting enzyme 2)
```

Searches for the genus and all species name belonging to the genus Nicotiana (+NT), e.g. *N. tabacum* and *N. benthamiana*. No old terms exist for Nicotiana.

© 2023 American Chemical Society. All rights reserved.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Nicotiana hierarchy in Medline

```
=> e Nicotiana+all/ct
 .  .  .
E143      1162          BT1   Solanaceae/CT
E144     36485           -->   Tobacco/CT
E145     36485           MN    B1.875.800.575.912.250.908.500.900./CT
                          RR    txid4085
                          RR    txid4097
                          DC    an INDEX MEDICUS major descriptor
                          NOTE  A plant genus of the family SOLANACEAE.
                                Members contain NICOTINE and other
                                biologically active chemicals; its dried
                                leaves are used for SMOKING.
                          INDX  as plant & for smoking ( = SMOKING), sniffing
                                ( = SNUFF see TOBACCO, SMOKELESS) & chewing (
                                = CHEWING TOBACCO see TOBACCO, SMOKELESS)
                          AQ    AE AH CH CL CY DE EM EN GD GE IM ME MI PH PO
                                PS RE TO UL VI
                          MHTH  NLM (1966)
E146         0            UF    Nicotiana/CT
E147         0            UF    Nicotiana tabacum/CT
E148         0            UF    Nicotiana tabacums/CT
E149         0            UF    Nicotianas/CT
E150         0            UF    Tobaccos/CT
E151         0            UF    tabacum, Nicotiana/CT
E152     28641            RT    Nicotine/CT
E153    149347            RT    Smoking/CT
E154     12462            RT    Tobacco Use Disorder/CT
```

MeSH typically show the trivial name and might not go to the species level: **tobacco** genus.

MeSH have extensive **notes** to explain the descriptor.

© 2023 American Chemical Society. All rights reserved.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Nicotiana hierarchy in Embase

```
=> e Nicotiana+all/ct

E179        314    BT9    Streptophyta/CT
E180        827     BT8    Embryophyta/CT
E181       2366      BT7    vascular plant/CT
E182        274       BT6    seed plant/CT
E183      12772        BT5    angiosperm/CT
E184        363         BT4    eudicot/CT
E185          6          BT3    asterid (plant)/CT
E186         28           BT2    Solanales/CT
E187       1335            BT1    Solanaceae/CT
E188        774              -->    Nicotiana/CT
                               TYPE   medical term
                               HNTE   Creation date 01 JAN 2012
E189       3743               NT1    Nicotiana benthamiana/CT
E190       1736               NT1    Nicotiana tabacum/CT
```

The Embase Emtree thesaurus has incorporated an extended version of Medline's medical subject headings. For the genus Nicotiana, the two most prominent species are indexed.

© 2023 American Chemical Society. All rights reserved.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Use Case 1: Find publications related to the DXR target of antibiotic-resistant pathogen *Acinetobacter baumanii*

```
=> index bioscience -uspatfull -uspat2 -ifiall -genbank

INDEX 'ADISCTI, ADISINSIGHT, ADISNEWS, AGRICOLA, ANABSTR, BIOSIS,...'
38 FILES IN THE FILE LIST IN STNINDEX

=> s (acinetobacter baumanni# or A baumanni#) and (dxr or 1-deoxy-D-xylulose-5-
phosphate reductoisomerase or 1.1.1.267) not p/dt
. . .
=> D rank
F1            7   CAPLUS
F2            5   EMBASE
F3            5   SCISEARCH
F4            3   BIOSIS
F5            3   MEDLINE
F6            3   REAXYSFILEBI
F7            1   CABA
F8            1   DDFU
F9            1   DRUGU
F10           1   TOXCENTER
F11           1*  ADISINSIGHT
```

Open relevant biomedical databases in the INDEX environment using the bioscience cluster

Send query. Search at least the full name and it's typical abbreviation. Rank results by file

We will perform the actual search in the top 5 databases. . .

© 2023 American Chemical Society. All rights reserved.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Use Case 1: Find publications related to the DXR target of antibiotic-resistant pathogen *Acinetobacter baumanii*

```
=> fil caplus Embase medline biosis scisearch
. . .
=> s (acinetobacter baumanni# or A baumanni#) and (dxr or 1-deoxy-D-xylulose-5-phosphate
reductoisomerase or 1.1.1.267) not p/dt
L2             7 FILE CAPLUS
L3             5 FILE EMBASE
L4             3 FILE MEDLINE
L5             3 FILE BIOSIS
L6             5 FILE SCISEARCH

TOTAL FOR ALL FILES
L7            23 (ACINETOBACTER BAUMANNI# OR A BAUMANNI#) AND (DXR OR 1-DEOXY-D-X
                 YLULOSE-5-PHOSPHATE REDUCTOISOMERASE OR 1.1.1.267) NOT P/DT
=> dup rem l7
L8            12 DUP REM L7 (11 DUPLICATES REMOVED)
                  ANSWERS '1-7' FROM FILE CAPLUS
                  ANSWERS '8-9' FROM FILE EMBASE
                  ANSWER '10' FROM FILE BIOSIS
                  ANSWERS '11-12' FROM FILE SCISEARCH
```

Open files acc. to priority

Send the search again

Duplicate remove eliminates most of the dupliactes

© 2023 American Chemical Society. All rights reserved.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Use Case 1: Find publications related to the DXR target of antibiotic-resistant pathogen *Acinetobacter baumanii*

```
=> d bib hit 1-tot
```

CAplus

```
TI   Corrigendum: Prioritizing multiple therapeutic targets in parallel using
     automated DNA-encoded library screening [Erratum to document cited in
     CA169:459705]
IT   Animal gene
     RL: BSU (Biological study, unclassified); BIOL (Biological study)
        (Dxr; prioritizing multiple therapeutic targets in parallel
        using automated DNA-encoded library screening (Erratum))
IT   Acinetobacter baumannii
```

EMBASE

```
TI   Analysis of drug binding pockets and repurposing opportunities for
     twelve essential enzymes of ESKAPE pathogens.
CT   Medical Descriptors:
     Acinetobacter baumannii
     amino acid sequence
CT   Drug Descriptors:
     1 deoxy d xylulose 5 phosphate reductoisomerase: EC, endogenous compound
```

Example hits from the remaining 4 databases

The CAplus text modifier lists the relevant target

Embase CTs names the species and target

© 2023 American Chemical Society. All rights reserved.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Use Case 2: Find publications related to a specific bacterial strain: Ab5075

```
=> index bioscience
. . .
=> s (Ab5075 or Ab 5075) not (Ab5075 or Ab 5075)/ti,ab
F1            24   BIOSIS
F2             5   EMBASE
. . .
=> fil biosis; s (Ab5075 or Ab 5075) not (Ab5075 or Ab 5075)/ti,ab
L30           24 (AB5075 OR AB 5075) NOT (AB5075 OR AB 5075)/TI,AB

=> d ti hit 1-3
TI   Rescuing Tetracycline Class Antibiotics for the Treatment of
     Multidrug-Resistant Acinetobacter baumannii Pulmonary Infection.
ORGN Classifier
        Neisseriaceae   06507
     Super Taxa
        Gram-Negative Aerobic Rods and Cocci; Eubacteria; Bacteria;
        Microorganisms
     Organism Name
        Acinetobacter baumannii (species): pathogen, strain-AB0057,
        strain-MS14413, strain-PGC-204089, strain-AB5075
     Taxa Notes
        Bacteria, Eubacteria, Microorganisms
```

Focus on the indexing for strains, i.e. where hits are not found in title and abstract. Biosis and Embase show the most answers.

Biosis indexing often mentions the strain in the organism field.

© 2023 American Chemical Society. All rights reserved.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society

Use case 3: Use thesauri to collect synonyms for a more comprehensive basic index text search

```
=> fil caplus
=> e danio rerio+all/ct

-->  Danio rerio/CT
  HNTE Valid heading during
       volume 126 (1997) to
       present.
  OLD  Brachydanio rerio/CT
  OLD  Zebra fish/CT
  UF   Brachydanio frankei/CT
  UF   MeSH ID: D015027/CT
  UF   Zebrafish/CT
  UF   Zebrafishes/CT
  NT1  Danio rerio albino/CT
  NT1  Danio rerio frankei/CT
```

```
=> fil medline
=> e danio rerio+all/ct

  -->   Danio rerio/CT
   USE   Zebrafish/CT

=> e zebrafish+all/ct
   -->   Zebrafish/CT
   . . .
  UF    B. rerio/CT
  UF    Brachydanio rerio/CT
  UF    D. rerio/CT
  UF    D. rerios/CT
  UF    Danio rerio/CT
  UF    Fishes, Zebra/CT
  UF    Zebra Fish/CT
  UF    Zebra Fishes/CT
  UF    Zebra danio/CT
  UF    Zebra danios/CT
  UF    Zebrafishes/CT
  UF    danio, Zebra/CT
```

```
=> fil embase
=> e zebrafish+all/ct

-->   zebrafish/CT
 TYPE  medical term
 NTYP  MeSH descriptor
 USE   zebra fish/CT

=> e zebra fish+all/ct

 -->   zebra fish/CT
  TYPE  medical term
  HNTE  Creation date 01 JAN 1984
  HNTE  loss of broader term relation: 2015-09-01 Old
        BT: Cypriniformes
  UF    Brachydanio rerio/CT
  UF    Cyprinus rerio/CT
  UF    Danio frankei/CT
  UF    Danio rerio/CT
  UF    zebra danio/CT
  UF    zebrafish/CT
  UF    zebrafish (MeSH)/CT
  NT1   Zebra fish AB strain/CT
  NT1   transgenic zebrafish/CT
```

© 2023 American Chemical Society. All rights reserved.

Summary

- Use EXPAND and database-specific CT hierarchies to look up scientific names used for indexing
- Search controlled terms for precise matches, use relationship codes to include further taxonomic levels
- Consider searching on a broader taxonomic level such a the genus or family
- Expand your search to the basic index when comprehensiveness is the target
- Look up CT hierarchies to create keyword queries including scientific names, their abbreviated form as well as product and lab jargon names

© 2023 American Chemical Society. All rights reserved.

Need support?

CAS Customer Center
1-800-753-4227
help@CAS.ORG

Monday 8:00 AM to Friday 12:00 midnight
U.S. Eastern Time

EMEA Customer Center
(+49) 721 95 883 155
EMEAhelp@CAS.org

Monday-Friday
8:30 AM to 5:00 PM CET

© 2023 American Chemical Society. All rights reserved.

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS
A division of the American Chemical Society