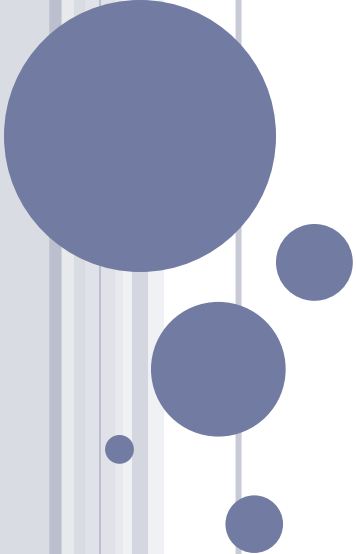




PARSING THE LEGACY DDDB

--- OLD BUT NEW

国立遺伝学研究所 DDBJ

遺伝子発現解析研究室

小笠原 理

目次

- DDBJのレガシー部分を一通り
- 生物種についてのDBと和名辞書



<http://www.insdc.org/>



INSDC International Nucleotide Sequence Database Collaboration

ABOUT INSDC POLICY ADVISORS DOCUMENTS



International Nucleotide Sequence Database Collaboration

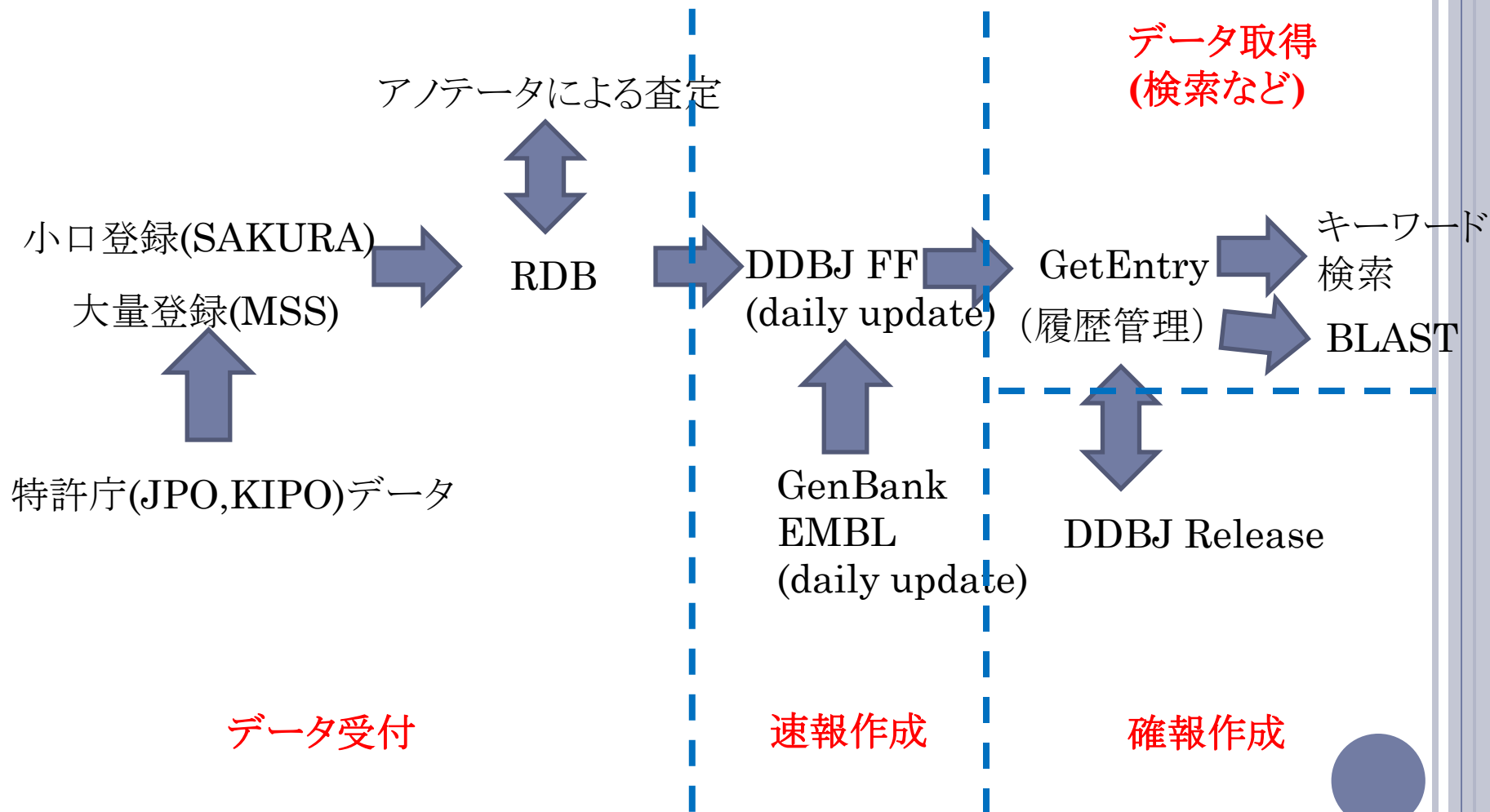
- The International Nucleotide Sequence Databases (INSD) have been developed and maintained collaboratively between [DDBJ](#), [ENA](#), and [GenBank](#) for over 18 years.
- The INSDC advisory board, the [International Advisory Committee](#), is made up of members of each of the databases' advisory bodies. At their most recent meeting, members of this committee unanimously endorsed and reaffirmed the existing data-sharing policy of the three databases that make up the INSDC, which is stated below.
- Individuals submitting data to the international sequence databases should be aware of [INSDC policy](#).

How to submit data

- For full details of how to submit data to the databases, please select a collaborating partner.
- [DDBJ](#), [ENA](#), [GenBank](#)
- The INSDC Feature Table Definition Document is available [here](#).

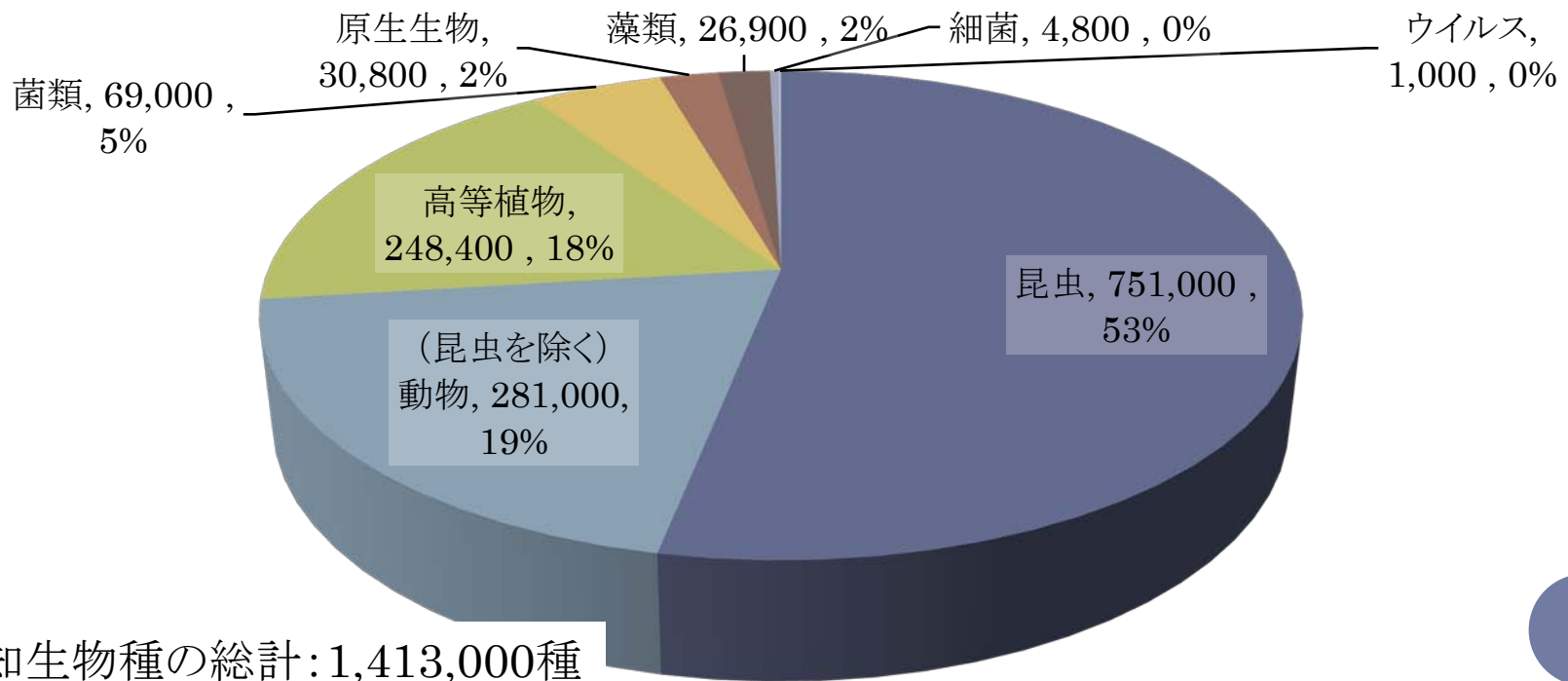
タスク: 研究者が決定したDNA塩基配列情報をデータベース化し、他の研究者が利用できるようにする。

DDBJにおけるデータ受付からリリース作成までの流れ



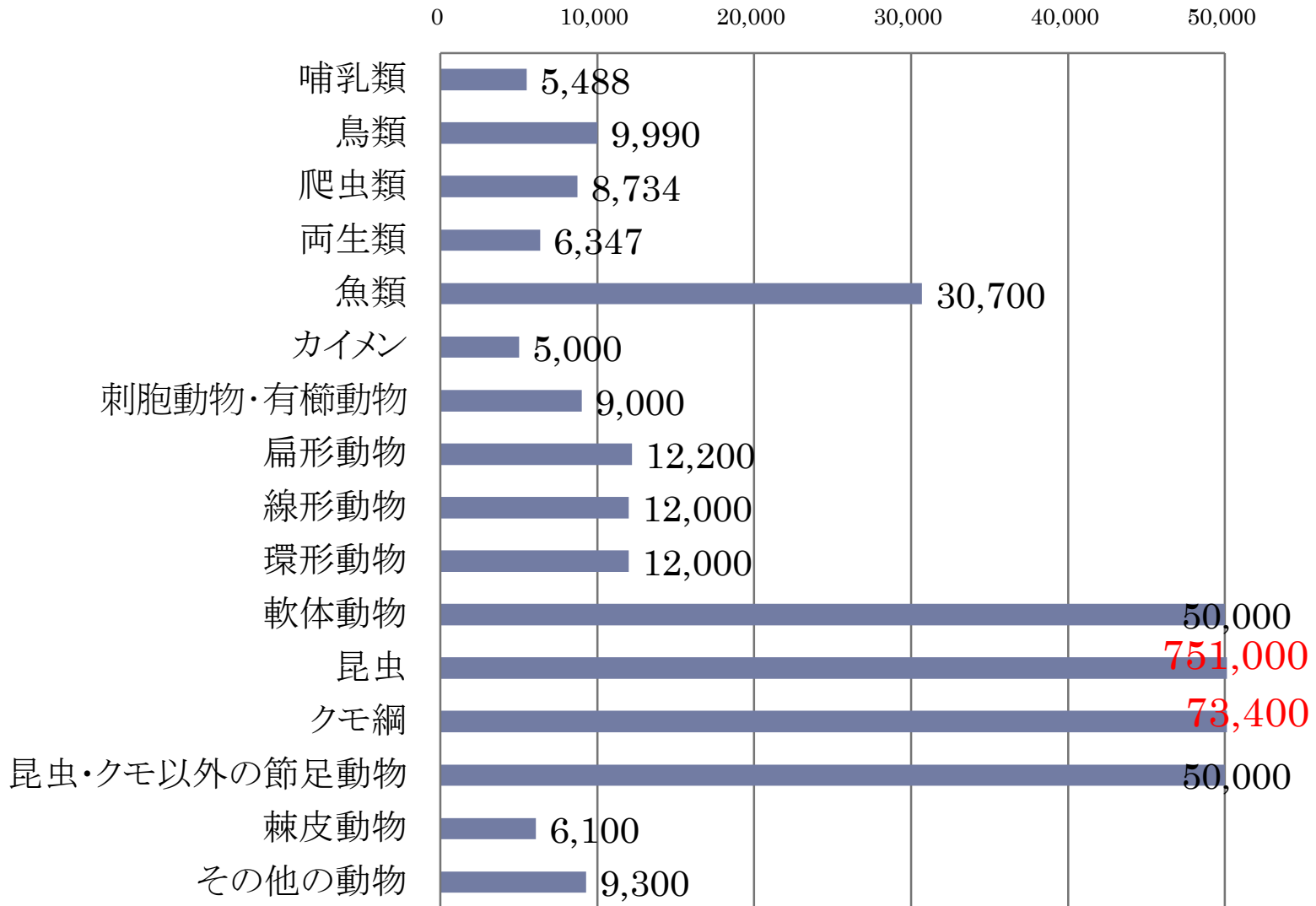
生物種数について

- 現存する生物種は5000万種～1億種と推定されている。
- 昆虫は3000万種、被子植物は1000万種とも言われる。
- 既知種(学名をもつ種)は、140万種～180万種。

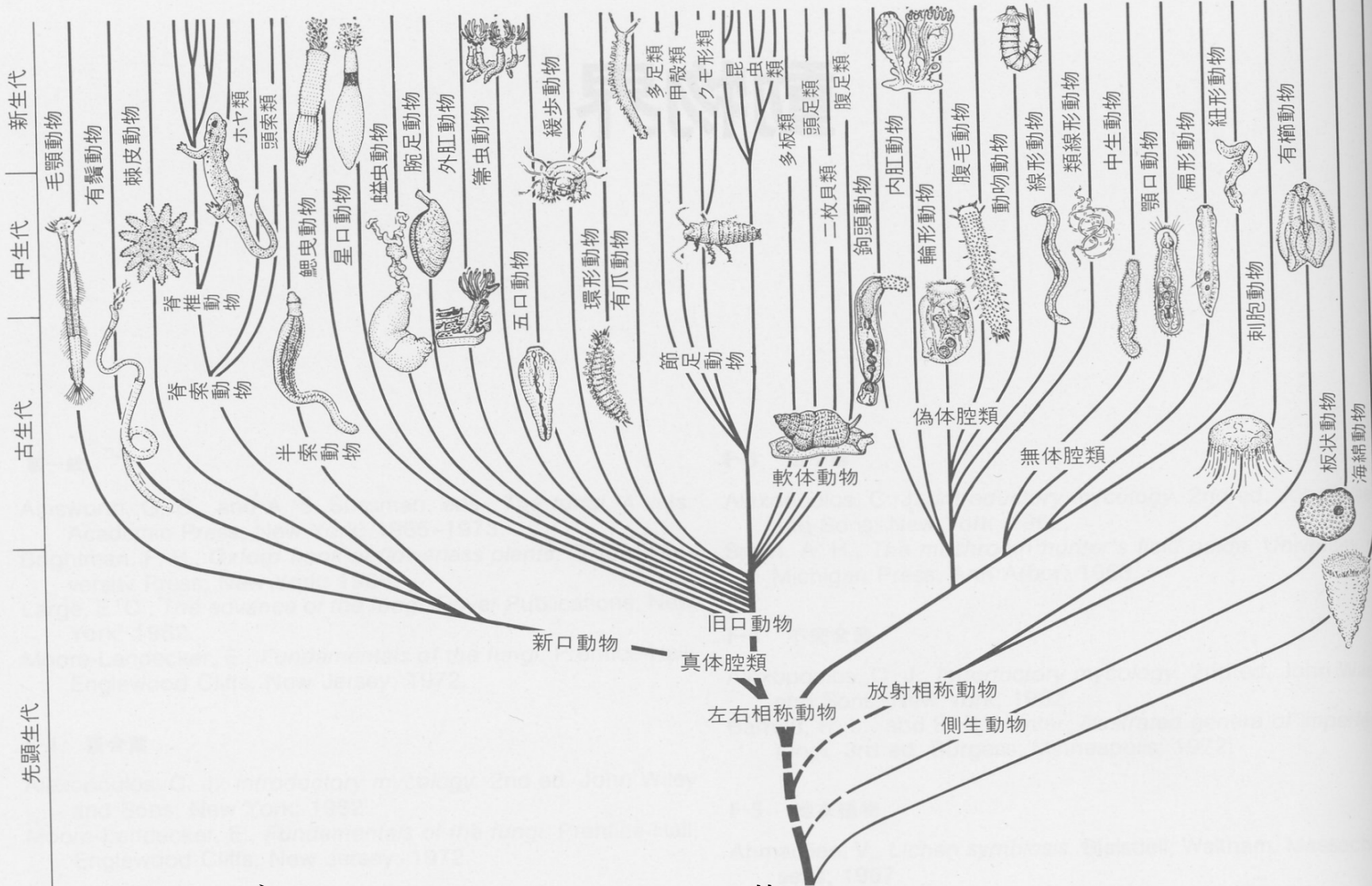


E.O.Wilson “The Diversity of Life” (1992) より。

既知の動物の種数

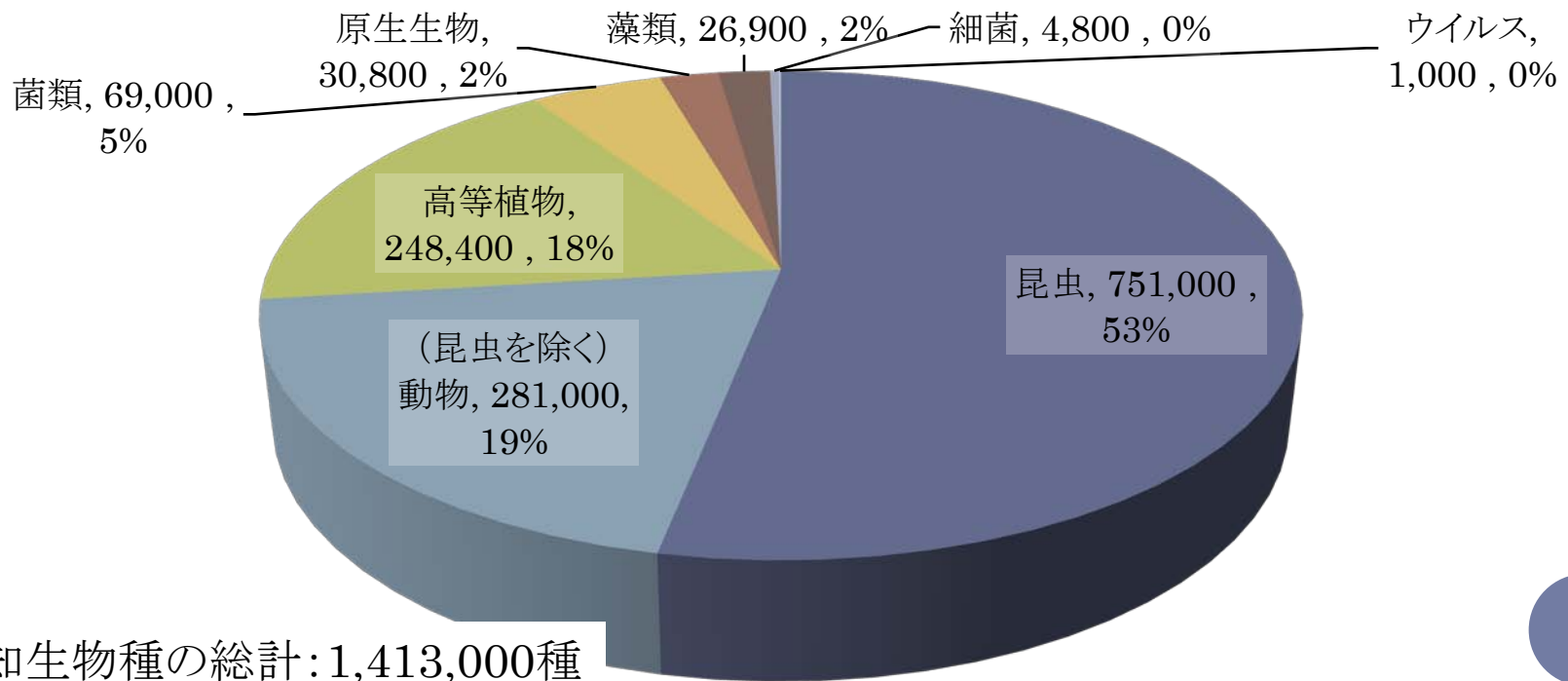


E.O.Wilson “The Diversity of Life” (1992) より改変



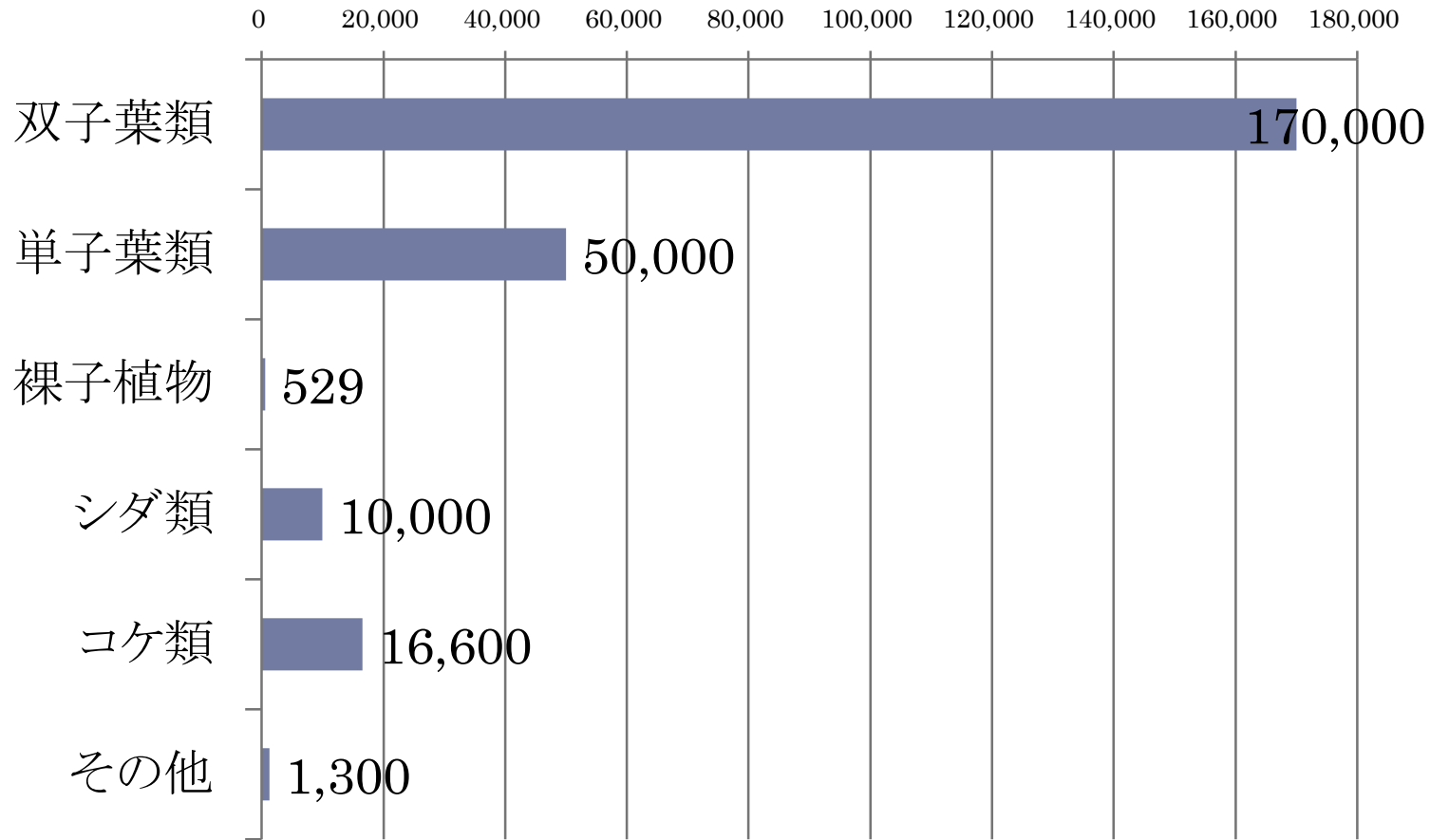
リン・マルグリス、カーリーン・V・シュバルツ著
 川島誠一郎・根平邦人訳
 「五つの王国 (Five Kingdoms)」日経サイエンス社 (1987)

- 現存する生物種は5000万種～1億種と推定されている。
- 昆虫は3000万種、被子植物は1000万種とも言われる。
- 既知種(学名をもつ種)は、140万種～180万種。



E.O.Wilson “The Diversity of Life” (1992) より。

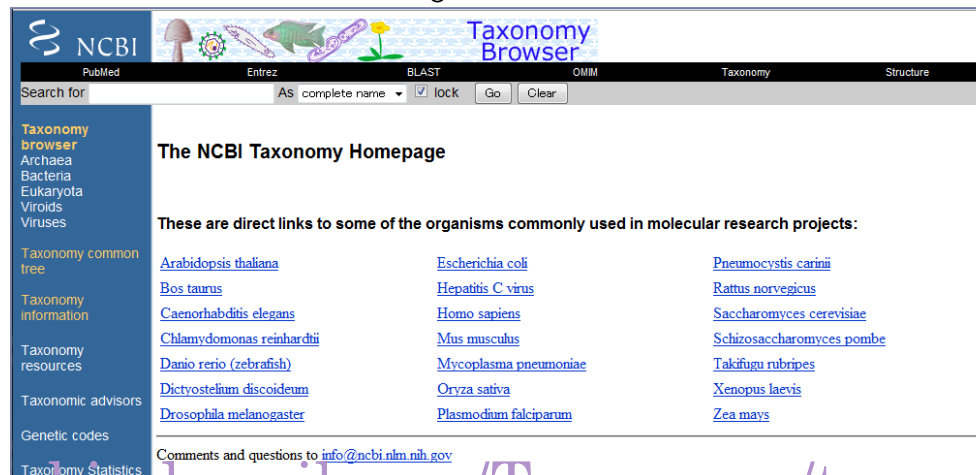
既知の高等植物の種数



E.O.Wilson “The Diversity of Life” (1992) より。

NCBI TAXONOMY

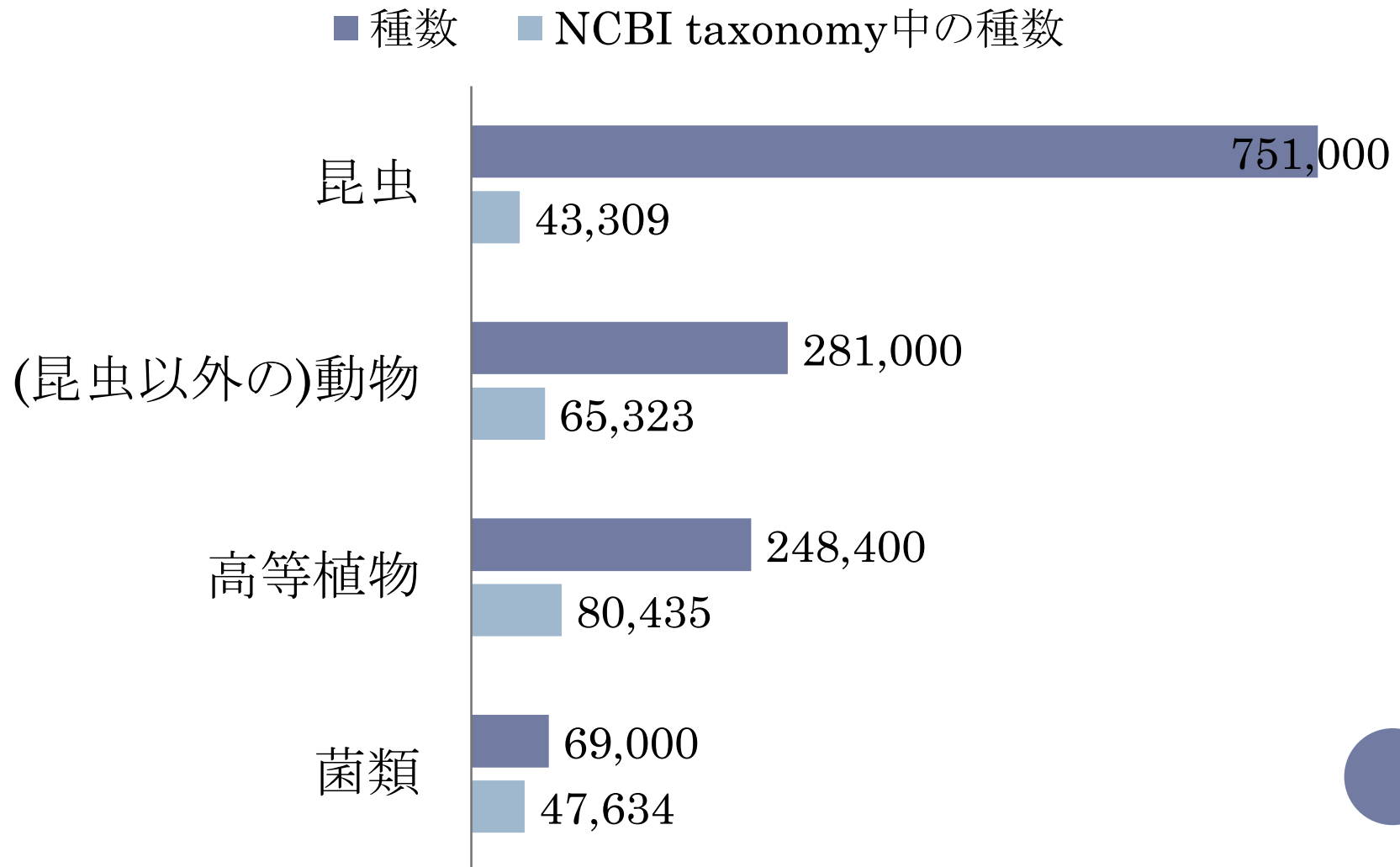
- DDBJ/EMBL/GenBankに登録されている塩基配列が由来する生物種名とその分類は、ここで管理されている。
- 登録されている生物種数は23万種* (既知生物種の約15%)
- 逆に、登録されている生物種のほとんどについて、なんらかの塩基配列がとられている。



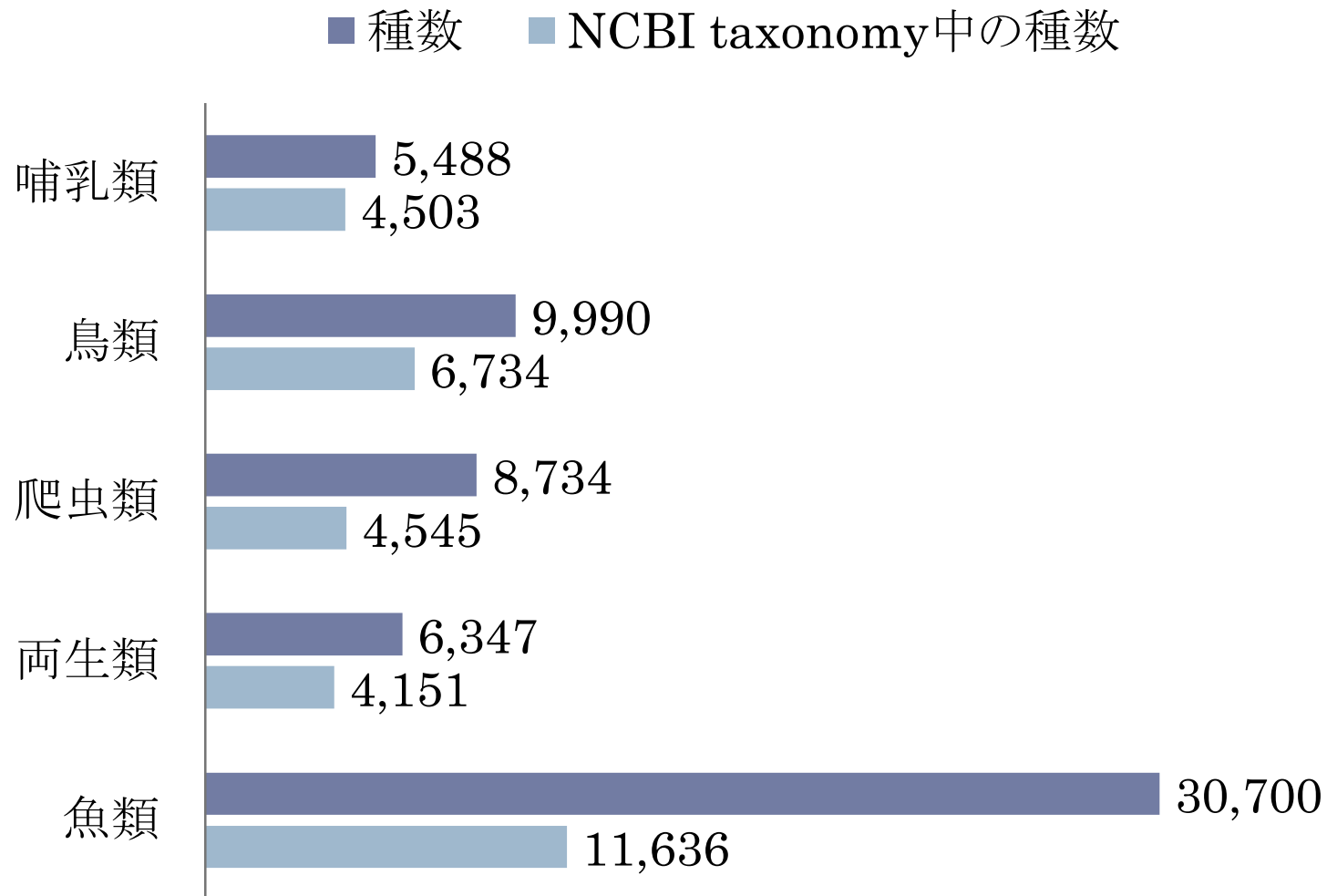
<http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/>

* NCBI Taxonomy Statisticsより。

NCBI TAXONOMYに登録されている種数 (全体)

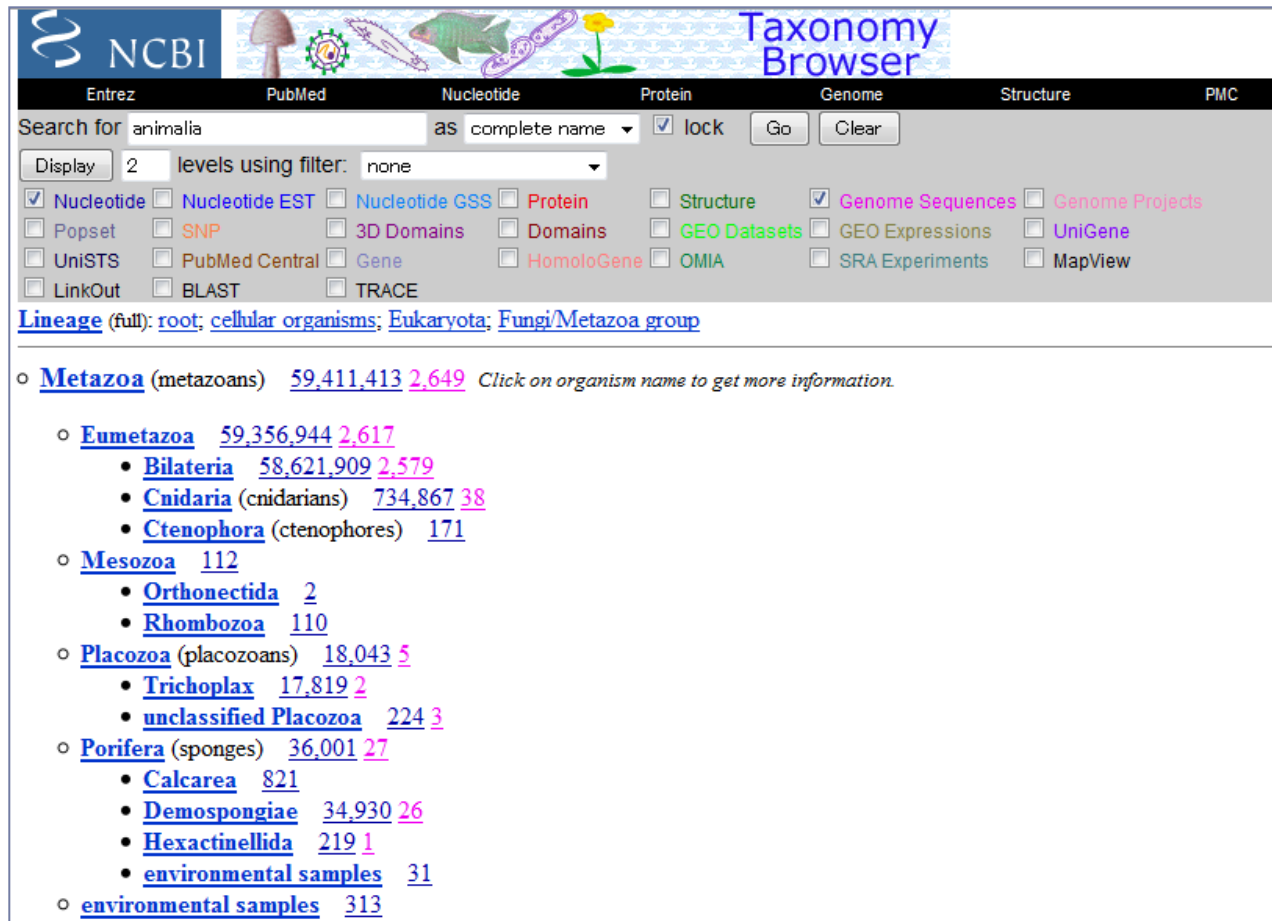


NCBI TAXONOMYに登録されている種数 (脊椎動物)



TAXONOMY BROWSERを使って種の属性情報を見る

- 例えば、とられた塩基配列の総数や読まれたゲノムの本数を表示できる。(表示してみてください)



The screenshot displays the NCBI Taxonomy Browser interface. At the top, the NCBI logo is on the left, and the title 'Taxonomy Browser' is on the right. Below the logo, there are navigation tabs: Entrez, PubMed, Nucleotide, Protein, Genome, Structure, and PMC. The search bar contains 'animalia' with a dropdown menu set to 'as complete name'. There are 'Go' and 'Clear' buttons. Below the search bar, there are options for 'Display' (set to 2) and 'levels using filter: none'. A grid of checkboxes allows users to select different data types: Nucleotide, Nucleotide EST, Nucleotide GSS, Protein, Structure, Genome Sequences, Genome Projects, Popset, SNP, 3D Domains, Domains, GEO Datasets, GEO Expressions, UniGene, UniSTS, PubMed Central, Gene, HomoloGene, OMIA, SRA Experiments, MapView, LinkOut, BLAST, and TRACE. Below the checkboxes, the 'Lineage (full)' is shown: root; cellular organisms; Eukaryota; Fungi/Metazoa group. The main content area shows a hierarchical list of taxa with associated counts. The 'Metazoa' (metazoans) entry is highlighted, showing 59,411,413 Nucleotide sequences and 2,649 Genome Sequences. Below it, a list of sub-taxa is shown, including Eumetazoa, Mesozoa, Placozoa, and Porifera, each with their respective counts.

NCBI Taxonomy Browser

Entrez PubMed Nucleotide Protein Genome Structure PMC

Search for animalia as complete name lock Go Clear

Display 2 levels using filter: none

☒ Nucleotide ☐ Nucleotide EST ☐ Nucleotide GSS ☐ Protein ☐ Structure ☒ Genome Sequences ☐ Genome Projects

☐ Popset ☐ SNP ☐ 3D Domains ☐ Domains ☐ GEO Datasets ☐ GEO Expressions ☐ UniGene

☐ UniSTS ☐ PubMed Central ☐ Gene ☐ HomoloGene ☐ OMIA ☐ SRA Experiments ☐ MapView

☐ LinkOut ☐ BLAST ☐ TRACE

Lineage (full): root; cellular organisms; Eukaryota; Fungi/Metazoa group

o **Metazoa** (metazoans) 59,411,413 2,649 Click on organism name to get more information.

- o **Eumetazoa** 59,356,944 2,617
 - **Bilateria** 58,621,909 2,579
 - **Cnidaria** (cnidarians) 734,867 38
 - **Ctenophora** (ctenophores) 171
- o **Mesozoa** 112
 - **Orthonectida** 2
 - **Rhombzoa** 110
- o **Placozoa** (placozoans) 18,043 5
 - **Trichoplax** 17,819 2
 - **unclassified Placozoa** 224 3
- o **Porifera** (sponges) 36,001 27
 - **Calcarea** 821
 - **Demospongiae** 34,930 26
 - **Hexactinellida** 219 1
 - **environmental samples** 31
- o **environmental samples** 313

生物種の和名の収集



Taxonomy Browser

Entrez PubMed Nucleotide Protein Genome Structure PMC

Search for animalia as

Display 2 levels using filter: none

☒ Nucleotide ☐ Nucleotide EST ☐ Nucleotide
☐ Popset ☐ SNP ☐ 3D Domain
☐ UniSTS ☐ PubMed Central ☐ Gene
☐ LinkOut ☐ BLAST ☐ TRACE

[Lineage \(full\):](#) [root](#); [cellular organisms](#); [Eukary](#)

- o [Metazoa](#) (metazoans) 59,411,413 2,64
 - o [Eumetazoa](#) 59,356,944 2,617
 - [Bilateria](#) 58,621,909 2,579
 - [Cnidaria](#) (cnidarians) 734,867
 - [Ctenophora](#) (ctenophores) 17
 - o [Mesozoa](#) 112
 - [Orthonectida](#) 2
 - [Rhombozoa](#) 110
 - o [Placozoa](#) (placozoans) 18,043 5
 - [Trichoplax](#) 17,819 2
 - [unclassified Placozoa](#) 224 3
 - o [Porifera](#) (sponges) 36,001 27
 - [Calcarea](#) 821
 - [Demospongiae](#) 34,930 26
 - [Hexactinellida](#) 219 1
 - [environmental samples](#) 31

Species Dictionary

root > Eukaryota > Fungi/Metazoa group > Metazoa (後生動物)

last data update: 2009.10.29

[Expand][Collapse]

Taxonomy	ID	Rank	Japanese	English
Metazoa	33208	kingdom	後生動物	metazoans, multicellular animals
1. environmental samples	212041	no rank		
1. Metazoa environmental sample	585499	species		
2. uncultured marine metazoan	329654	species		
3. uncultured metazoan	175278	species		
2. Eumetazoa	6072	no rank	真正後生動物 亜界	
1. Bilateria	33213	no rank	左右相称動物	
2. Cnidaria	6073	phylum	刺胞動物門	cnidarians, coelenterates
3. Ctenophora	10197	phylum	有櫛動物門	coelenterates, comb jellies, ctenophores,

http://bm2.genes.nig.ac.jp/Integrated_BodyMap/species.php

和名辞書のページ(DBCLS)

ライフサイエンス統合データベースプロジェクト-統合ホームページ-LSDB - Mozilla Firefox

ファイル(E) 編集(E) 表示(V) 履歴(S) ブックマーク(B) ツール(I) ヘルプ(H)

http://lifesciencedb.jp/

よく見るページ Firefox を使いこなそう 最新ニュース OgaNote projects info eljiro Google Gmail

ライフサイエンス統合データベー... x ライフサイエンス統合データベー... x Gmail - 検索結果 - osamu.ogas... x ジャワマメジカ - Google 検索 x Species Dictionary x

検索

生命科学データベース 横断検索

蛋白質核酸酵素 全文検索

文科学「ゲノム」研究報告書 全文検索

TogoProt (蛋白質関連データベース 統合検索)

new

OReFil (オンラインリソースファインダー)

Allie (略語の正式名称を検索)

inMeXes (文献中の英語表現を軽快に検索)

データベース

DNAデータベース 総覧と検索 (DDBJ/EMBL/GenBank)

遺伝子発現バンク(GEO)目次

KazusaAnnotation & Navigation (かずさDNA研究所)

KazusaMart (かずさDNA研究所)

ゲノムネット 医薬品データベース (京大)

Wired-Marker

MIGAP (微生物ゲノムアノテーションパイプライン)

DBCLS Galaxy

基盤技術開発

TogoDB (誰でもデータベースが構築できる)

TogoWS (ウェブサービスの標準化)

OpenID 認証システム

統合DB情報基盤プロジェクト (DBCLS)

辞書の構築と公開

LSDB Lab

BioHackathon(DBCLS バイオハッカソン) 2010, 2009, 2008

教材・人材育成

統合TV (DBやツールの動画教材)

MotDB (教育・人材育成のサイト)

を中心にご覧 2010-08-06 (Fri) 19:50:35

統合データベース講演会 AJACSみちのく開催 2010-08-05 (Thu) 11:12:50

ニュース

ipsから生殖細胞 慶大など初の研究着手へ

ipsから生殖細胞 研究へ 不妊治療に応用期待情報公開求める声も

ミラジエロの「隠し絵」神の姿に脳幹解剖図

バナーリンク

Anatomography BodyParts3D

BodyParts3D/Anatomography 使い倒し系チャンネル

統合TV

統合TV

OReFil

Online Resource Finder for Lifesciences

OReFil

完了

http://lifesciencedb.jp/

和名辞書のページ(DBCLS)



辞書

Guest Account | アカウント | ログイン | English

ホーム

データベース

検索

ツール

ダウンロード

About us

辞書

辞書の構築と公開

データベースの多くは遺伝子や蛋白質のIDや、文献IDなどのリンクが存在し、その情報を用いれば相互に“統合”することが可能です。場合によっては配列情報を使うことも可能です。一方、文献や注釈情報など自由文で書かれたもの、あるいは遺伝子や生物の名前など別名があるものは、辞書やシソーラス無しでの整理・統合は困難です。統合データベースプロジェクトでは、日英の翻訳情報がそれぞれの出典においてどのように記載されているかを記述したメタ用語集や、基本概念での関連付けや整理を可能にするための用語シソーラ作成のおよび用語の構造化(いわゆるオントロジー)を進めています。

DBCLSメタ用語集

メタ用語集 学術用語編

- [概要版](#)
- [ダウンロード](#) (zip形式 1.3M)

メタ用語集 学名編

- [概要版](#)
- [ダウンロード](#) (zip形式 8.6M)

解説

- [生物学辞書の構築](#) (詳しい解説です)
- [デジタル素材の充実を-メタ学術用語の公開\(LSDB Social Action\)](#)
- [日本語による質問で世界の生命科学情報を検索するために-デジタル対訳辞書の作成](#) (統合DBシンポジウム ポスター発表)

利用サイト

- [Species Dictionary](#) - 学名和名検索
- [地球規模生物多様性情報機構](#) 日本ポータルサイト 検索
- [DNAデータベース総覧と検索における生物種名の和訳表示](#)

和名の収集(1) データソース

- 今泉吉典 世界哺乳類和名辞典 平凡社 (2005)
- 山階芳麿 世界鳥類和名辞典 大学書林(1993)
- 中坊徹次 日本産 魚類検索 全種の同定 第二版 東海大学出版会 (2000)
- 矢田 脩,森本 桂,平嶋義宏 新訂 原色昆虫大圖鑑 第 I,II,III 卷 北隆館 (2007)
- 岡田 要 復刻版 新日本動物圖鑑 上中下 北隆 (2004)
- 奥谷喬司 日本近海産貝類図鑑 東海大学出版会 (2000)
- 牧野 富太郎 新牧野日本植物圖鑑北隆館 (2008)
- 等々 ...



和名の収集(2) 種の数

