

Integbioデータベースカタログ

– DB情報の活用を目指して

トーゴーの日シンポジウム2020

○信定知江¹⁾、建石由佳¹⁾、坂東明日佳¹⁾、宮崎敦子¹⁾、酒井智子¹⁾、栗原英輔²⁾、加藤健弘²⁾

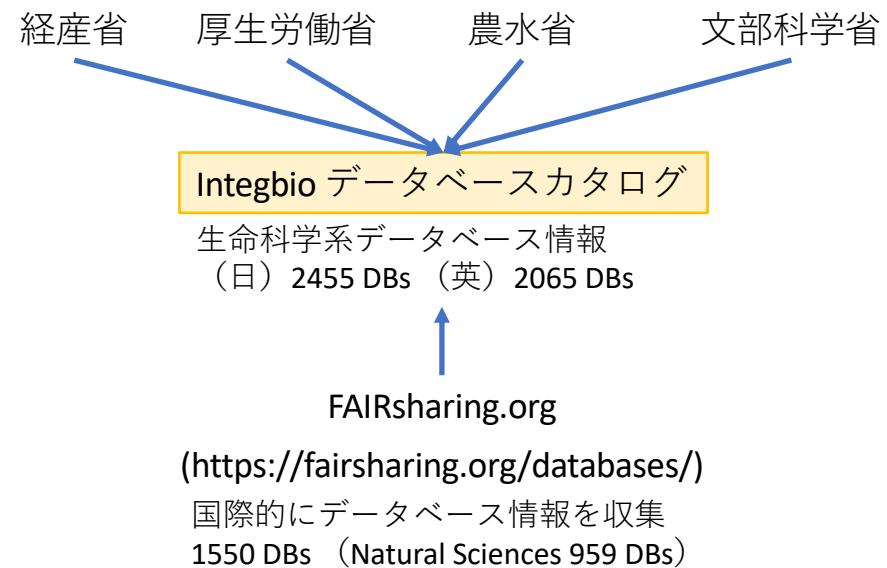
1) 国立研究開発法人 科学技術振興機構 バイオサイエンスデータベースセンター

2) 株式会社日立製作所

Integbioデータベースカタログ (<https://integbio.jp/dbcatalog/>) は文部科学省・厚生労働省・経済産業省・農林水産省の連携のもと、主に国内で作成された生命科学系データベース（DB）の所在と概要の提供を目的として始まったサービスである。現在では1300件近くの国内DBに加え、国外主要DB、FAIRsharing.org (<https://fairsharing.org/>) 由来やNucleic Acid Research (NAR) Database Issue由来のDBなど、計2450件を超えるDB情報を公開している。近年は主に国内に留まらないDB情報収集経路の確立や、種々DB情報の拡充を行ってきた。

国内外の主要なDBをキャッチアップする道筋ができた中、収録DB情報の活用が次の課題となる。羅列的にDB情報を提供するだけでなく、DB情報を整理し、意味のある関係を示すことで、目的DBへのスムーズなリードが期待できる。

本ポスターでは、現在考え得るデータベース情報の付加的な提供方法を3つ挙げた。それについてニーズ、問題点等を議論したい。



Integbioデータベースカタログ

- DB情報の活用を目指して

トーゴーの日シンポジウム2020

トップページ

- 生命科学系データベースを一覧から探す -

Integbioデータベースカタログ

全条件をリセット

一覧内を検索する

一覧を絞り込む

生物種

- + 動物 (1013)
- + 植物 (400)
- + 原生生物 (92)
- + 菌類 (183)
- + 真正細菌 (252)
- 古細菌 (81)
- ウイルス (93)

キーワード検索

DB 絞り込み

- 生物種
- タグ (対象)
- タグ (データの種類)
- 稼働状況
- 地域
- その他の条件
 - LSDBアーカイブ収載のDB
 - 統合TVで紹介のあるDB
 - 一括ダウンロード可のDB

データベースのレコード一覧 (件数: 2455)

最初へ 前へ 1 2 3 4 5 6 7 8 9 10 次へ 最後へ

並べ替え: レコード公開順

Vesiclepedia
運用機関: La Trobe Institute for Molecular Sciences
生物種: All
説明: 細胞外小胞 (Extracellular vesicle : EV) のデータベースです。生成様式によってエクトソーム、エキソソーム、アポトーシス小体の3種類に分類される細胞外小胞について、内容物 (タンパク... 詳細へ)

Stemformatics
運用機関: University of Melbourne
生物種: Homo sapiens | Mus musculus
説明: 幹細胞に関する遺伝子発現のデータベースです。本データベースのデータは、マイクロアレイ、RNA sequencing、single cell profiling法など遺伝子発現を調べた公共の生データから一連の品質管... 詳細へ

OrthoInspector
運用機関: iCube, Université de Strasbourg
生物種: Eukaryota | Archaea | Bacteria
説明: 3つの生物界 (真核生物、古細菌、バクテリア) にわたるタンパク質のオーソログのデータベースです。タンパク質のオントロジー、配列、タンパク質ドメイン、オーソログなどを収録しています... 詳細へ

MoonDB
運用機関: TAGC, Inserm U1090
生物種: Homo sapiens | Mus musculus | Drosophila melanogaster | Caenorhabditis elegans | Saccharomyces cerevisiae
説明: 相関関係のない複数の機能を持つ多機能タンパク質やMoonlightingタンパク質のデータベースです。手動でキュレート、または推測された多機能タンパク質について、遺伝子名、機能、UniProtKB... 詳細へ

レコード詳細ページ

レコード詳細

KNAPSAck Biological Activity (Metabolite Activity)

ID: NBDC01175
名称: KNAPSAck Biological Activity (Metabolite Activity)
別称:
URL: http://www.knapsackfamily.com/MetaboliteActivity/top.jsp
運用機関名: 奈良先端科学技術大学院大学 情報科学研究科 (J-GLOBAL)へのリンク
運用機関所在国・地域: 日本
説明: 代謝物質とそれらの生物学的性質の関係に関するデータベースです。代謝物質名、生物学性(機能)、タグ(対象)となる生物種名および参考文献の情報を収録しています。各データは代謝物質名、C_ID、CAS_ID、生物活性、生物種、分子式で検索することが可能です。
KNAPSAck Family (http://integbio.jp/databatalog/record?id=01086) が1つです。

生物種:
タグ (対象):
タグ (データの種類):
論文等 (PubMed ID/DOI):
言語:
稼働状況:
一括ダウンロードデータへのリンク:
一括ダウンロードデータのデータ範囲:
データベースの利用許諾へのリンク:
データベースの使い方:
データベースの連絡先:
API / SPARQL endpoint:
LSDBアーカイブへのリンク:
MEDBISデータベース便覧へのリンク:
統合TVへのリンク:
FAIRsharingへのリンク:
似似データベース:
レコード管理者:
レコードの由来:
レコード公開日: 2012-07-10
レコード最終更新日: 2019-10-20
本レコードの利用条件: Creative Commons CC0 license

Integbio Database Catalog

BY

Licensed under a Creative Commons 表示4.0国際ライセンス
©2020バイオサイエンスデータベースセンター

Integbioデータベースカタログ

- DB情報の活用を目指して

トゴーの日シンポジウム2020

(構想1) 横断検索との連携によるデータベース内のデータ検索

DB情報

レコード詳細

KNAPSAck Biological Activity (Metabolite Activity)

ID: KNAPSAck Biological Activity (Metabolite Activity)
名前:
別称:
URL:
運用機関名:
運用機関所在国・地域:
説明:
生物種:
タグ (対象):
タグ (データの種類):
論文等 (PubMed ID/DOI):
PubMed ID: 22123792
DOI: -
PubMed ID: 23509110
DOI: -

一括ダウンロードデータのデータのリンク:
データベースの利用許諾へのリンク:
データベースの使い方:
データベースの連絡先:
API / SPARQL endpoint:
LSDBアーカイブへのリンク:
MELALSデータベース便覧へのリンク:
統合TVへのリンク:
FAIRsharingへのリンク:
類似データベース:
KNAPSAck Twins (化合物類似度データベース)
Metabolite Ecology DB
KNAPSAck Biological Activity (Natural Activity)
Flavonoid Database
蛋白質データベース
Bio-MassBank
KNAPSAck
Plantmetabolomics.org
NPedia (Natural Products Encyclopedia) 化合物検索

レコード管理者:
レコードの由来:
レコード公開日:
レコード最終更新日:
本レコードの利用条件: Creative Commons CC0 license



DB収録データ

Integbioデータベースカタログ

検索

検索対象: KNAPSAck Biological Activity
検索キーワード: kinase 検索

全 4 件 (1 - 4 件目) (20 件ずつ表示)

生物活性 (代謝物) データ: 5487 [KNAPSAck Biological Activity (Metabolite Activity)]
http://togodb.biosciencedbc.jp/knapsack_metaboliteactivity_main/show/5487
C00000674 luteolin inhibit protein kinase C - Harborne,Phytochemical Dictionary Second Edition,Taylor and Francis,(1999),Chapter37 ...

生物活性 (代謝物) データ: 2606 [KNAPSAck Biological Activity (Metabolite Activity)]
http://togodb.biosciencedbc.jp/knapsack_metaboliteactivity_main/show/2606
C00001000 sanggenon C inhibit activation of protein kinase C in a dose-dependent manner -

生物活性 (代謝物) データ: 2614 [KNAPSAck Biological Activity (Metabolite Activity)]
http://togodb.biosciencedbc.jp/knapsack_metaboliteactivity_main/show/2614
C00001001 sanggenon D inhibit activation of protein kinase C - Harborne,Phytochemical Dictionary Second Edition,Taylor and Francis,(1999),Chapter36 ...

生物活性 (代謝物) データ: 5439 [KNAPSAck Biological Activity (Metabolite Activity)]
http://togodb.biosciencedbc.jp/knapsack_metaboliteactivity_main/show/5439
C00004631 quercetin inhibit many enzymes, e.g., protein kinase C, lipogenases, lens aldos

Licensed under a Creative Commons 表示4.0国際ライセンス
©2020バイオサイエンスデータベースセンター



Integbioデータベースカタログ

– DB情報の活用を目指して

トーゴーの日シンポジウム2020

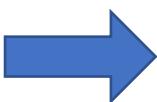
(構想 2) 主要DBのIDを持つDBリストによる絞り込み

レコード詳細

KNApSAcK Biological Activity (Metabolite Activity)

ID: NBDC01175
名前: KNApSAcK Biological Activity (Metabolite Activity)
別称: —
URL: http://www.knapsackfamily.com/MetaboliteActivity/top.jsp
運用機関名: 研究先端科学技術大学院大学 情報科学研究科 (J-GLOBALへのリンク)
運用機関所在国・地域: 日本
説明: 代謝物質とそれらの生物活性の関係に関するデータベースです。代謝物質名、生物活性(作用機序)、分子構造による検索機能、及び参考文献情報を収集しております。各データは、CAS登録番号、C、CA、生物活性、生物種、分子式で検索する事が可能です。
KNApSAcK Family (http://integbio.jp/database/records/00001086) の1つです。
生物種: —
タグ (対象): 代謝物
タグ (データの種類): 遺伝子動画、化学会議
PubMed ID/DOI: PubMed ID: 22123792
DOI: 23509110
言語: 英語
掲載状況: 理論中
—既ダウンロードデータへのリンク: https://dbarchive.bioclassedbc.jp/p/knapsack-metaboliteactivity/download.html
—既ダウンロードデータのデータベースへのリンク: http://pbioclassedbc.jp/archive/knapsack-metaboliteactivity/LATEST/README.html
データベースの利用許諾へのリンク: https://dbarchive.bioclassedbc.jp/p/knapsack-metaboliteactivity/lic.html
データベースの使い方: —
データベースの販路先: —
API / SPARQL endpoint: —
LSDBアーカイブへのリンク: http://dbarchive.bioclassedbc.jp/p/knapsack-metaboliteactivity/desc.html
MEDALSデータベース使用へのリンク: —
統合TVへのリンク: —
FAIRsharingへのリンク: KNApSAcK Twins (化合物類似度データベース)
ReSpect
Metabolite Ecology DB
KNApSAcK Biological Activity (Natural Activity)
Flavonoid Database
食品メタボロームプロジェクト
Bio-MassBank
NPDB
Plantmetabolomics.org
NPedia (Natural Products Encyclopedia) 化合物検索
類似データベース:
Integbio Database Catalog
CC BY
レコード管理者: —
レコードの由来: CC BY
レコード公開日: 2012-07-10
レコード最終更新日: 2019-10-20
レコードの利用条件: Creative Commons CC0 license

NCBI gene ID
UniProt ID
ChEBI
HGNC ID
 TAIR Locus_ID



TAIR Locus_ID (Atxgxxxxxx)をデータにもつDBs

NBDC00697 TAIR

ゲノム・遺伝子 (シロイスナズナ)

NBDC01667 Tomato Functional Genomics database

ゲノム・遺伝子 (トマト)

NBDC02447 Mum GARDEN

ゲノム・遺伝子 (キク)

NBDC01677 Protein Localization Database for Human and Arabidopsis

細胞内局在性

NBDC00374 ATTED-II

遺伝子発現

NBDC01041 PlantPromoterDB

転写開始点

NBDC00690 STRING

タンパク質相互作用

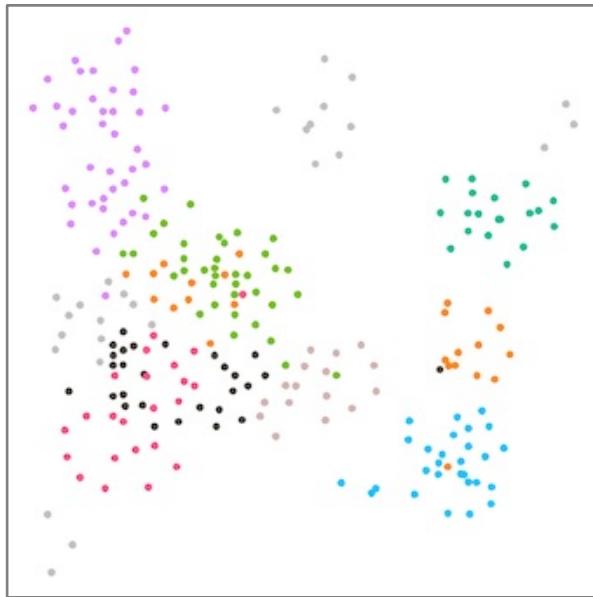
NBDC00118 KEGG Pathway

代謝ネットワーク

•

•

(構想3)類似データベースで絞り込む



タグ「遺伝的多様性」で絞り込み (243 DBs)

↓
タグと説明文によるクラスタリング (11 clusters)

cluster1

変異体 遺伝子多型 臨床
遺伝子
遺伝子
大規模 遺伝子 変異
規則
NCBI
表現型
科学 報告
挿入
Ontology
バイオリソース

cluster2

オントロジー/用語/命名法
クローニング
表現型
バイオリソース
センサー cDNA
長さ
マーカー 表現型

cluster3

mRNA マップ
cDNA/EST
ブタ 物理
BLAST EST
条件 植物
マーカー プライマー
cDNA 遺伝子地図

cluster4

臨床試験 薬剤
3D構造
細胞/オルガネラ
生物学
タンパク質 Target
分析
細胞
RNA
変異体
表現型
相互作用

cluster5

手法 アーカイブ
cDNA/EST 表現型
変異体 *リボジトリ*
タイプ *RNA* 表現型
突然変異
表現型 自己アレイ

cluster6

マーカー “画像/動画” 環境
長さ SSR *環境*
分離 分布 *表現型*
突然変異 分布 イヌ
BLAST 集約 遺伝子連鎖
マーカー ラット

cluster7

集約 遺伝子多型
診断 表現型
抽出 リスク
産業 バリアント
評価 個人 臨床
個人 遺伝子多型

cluster8

結合部位 “画像/動画”
cDNA/EST アミノ酸
細胞 RNA PTM
プロモーター相互作用
タンパク質
変異 翻訳後修飾
遺伝子多型 部分
変異 遺伝子多型

cluster9

リポジトリ 転写因子
結合部位 探索
挿入 検証 保存
改良 置換
プライマー

cluster10

アレイ 探索
連鎖 調節組み合わせ
ニワトリ 転写因子
遺伝子多型

cluster11

ハプロタイプ 医学
JSNP 探索
生命科学 生命科学
科学技術 検体

cluster1と5：説明文の「表現型」が同じ程度の寄与だが、タグで分けられる
cluster3と5：タグ「cDNA/EST」が同じ程度の寄与だが、他のタグと説明文で分けられる

黄色系統：タグ由来
青一緑系統：説明文由来

Integbioデータベースカタログ

– DB情報の活用を目指して

トーゴーの日シンポジウム2020

(構想3)類似データベースで絞り込む

cluster3
(28 DBs)



DB名：キーフレーズ（手動抽出）

Ensembl Genomes: ゲノムに関する情報/ゲノムアッセンブリー、遺伝子アノテーション、比較ゲノミクス、変異の情報

wFleaBase: ゲノムデータベース/ST、GSS、マイクロサテライトのライブラリー、遺伝子についての配列、染色体位置、相同遺伝子、タンパク質、オントロジー、発現情報

TOMATOMICS: トマトのEST配列、マイクロトムの完全長cDNA配列、iTAG2.4遺伝子モデルの配列

Striga hermonthica EST Database: ストライガ (*Striga hermonthica*) のEST情報/遺伝子名、BLAST、pSSR、pSNP名、GOアノテーション

Cassava Online Archive: 有用熱帯作物キャッサバのmRNA配列、EST、それらのアノテーション情報

Raphanus sativus Genome DataBase: ダイコン (*Raphanus sativus*、ラファナス サチパス) のDNAマーカーデータベース/EST-SSR、Isozyme marker、Morphological marker、R-EST、RAPD、RFLP、SSR、STS

Brown planthopper Maps & Markers Database: トビイロウンカ (BPH, *Nilaparvata lugens*) の遺伝子連鎖地図やマーカー

KaTomicsDB: トマトのゲノムデータベース/DNAマーカーやSNPのアノテーション（機能性マーカー）、Selected BAC clone Mixture (SBM)

Soybean Marker Database: ダイズのマーカー情報/EST-SSR/プライマー配列やEST配列

SNPSTR: マイクロサテライト(STR)マーカー/SNPSTRと呼ばれる複合遺伝マーカー

cluster5
(24 DBs)



IGDB.NSCLC: 肺がんに関わるゲノムセット/肺がん遺伝子やmicroRNA/マイクロアレイによる発現

PMutData: アミノ酸変異体の病理学的影響の予測データ/リポジトリ

STOREDB: 放射線生物学および放射線疫学分野のデータとリソースの共有/データをアーカイブ

JGA: 個人レベルでの遺伝的データと匿名化された表現型データを永続的にアーカイブし、共有するためのサービス

GnpIS: 多種統合的情報システム/遺伝的情報（遺伝地図、QTL、マーカー、多型、表現型、遺伝子型）とゲノムデータ（配列、物理的マップ、ゲノムアノテーション、発現データ）

openSNP: ゲノムデータ/SNPのジェノタイピングデータ、エクソンの配列決定データ、表現型に付されたアノテーション、定量化された自己追跡データ

The ribosomal RNA operon copy number database: 16S や 23SのリボソームRNAをコードする遺伝子のコピー数と生物種、NCBI BioSample ID、NCBI BioProject ID

EVA: 遺伝的変異タイプを対象としたオープンアクセスのアーカイブ

MMRRC: 遺伝子改变マウス系統およびマウスES細胞株を配布/変異遺伝子の位置や種類、表現型、コロニーの状態

QTL Archive: QTL（量的形質遺伝子座）研究から得られた生データへのアクセスを提供するアーカイブ