

# 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 の取り組み

2015.07.03

JaLC研究データDOI登録実験 中間報告会

櫛田達矢 (NBDC/JST)

# 目次

1. テスト実施方針
2. 生命科学系データベースアーカイブとは
3. クリエイティブ・コモンズ (CC)
4. テストの方法
  - 登録対象データと登録対象件数
  - テスト登録方法
5. 結果
6. 課題と検討事項
7. 本番登録の予定

# 1. テスト実施方針

- NBDCがサービスを提供している「生命科学系データベースアーカイブ」  
(<https://dbarchive.biosciencedbc.jp/>)  
に収録されているデータセットに対してDOIを付与するテストを行う。JaLCのDOI登録インターフェースを使用し、生命科学系データベースアーカイブのメタデータがDOI登録にもそのまま活用できるか、不足の項目があるかなどを確認し、本格運用開始に備えることを目的とする。

# 2. 生命科学系データベースアーカイブ

データセットを**長期安定に維持保管し、データ説明（メタデータ）を統一して検索**を容易にすると共に、**利用許諾条件などの明示**を行うことで、多くの人が容易にデータへアクセスしダウンロードを行えるようにするサービス



～あのデータベースが、丸ごとダウンロード可能に！～  
生命科学系データベース アーカイブ

アーカイブ内を横断検索

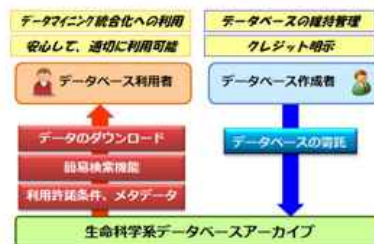
同人性検索  画像検索

ホーム アーカイブの説明 寄託応募要領 更新履歴 利用状況 ヘルプ お問い合わせ

いくら良質なデータベースでも、説明が十分でない、利用条件が明確でない、ダウンロードできないなどの理由で十分に利用され、引用され、相応しい評価をうける機会を逃していることがあります。

生命科学系データベースアーカイブは、国内のライフサイエンス研究者が生み出したデータセットをわが国の公共財としてまとめて長期安定に維持保管し、データ説明（メタデータ）を統一して検索を容易にすると共に、利用許諾条件などの明示を行うことで、多くの人が容易にデータへアクセスしダウンロードを行えるようにするサービスです（詳細説明）。

データを長期にわたり保全し、データベース作成者のクレジットを明示する一方、公的機関や民間等様々なユーザが利用しやすい形にすることで、それぞれの研究の生命科学へのいっそうの貢献を支援します。データベースの寄託を随時募集しています（寄託応募要領）。



**新アーカイブ情報**

2015/05/29 「NBDC NikkajiRDF」  
（国立研究開発法人 科学技術振興機構）を追加しました

2015/05/12 「RGP caps」（農業生物資源研究所）を追加しました

2015/04/02 「RGP estmap2001」（農業生物資源研究所）を追加しました

## アーカイブデータベース一覧（ヘルプ）

一覧内検索

全 102 件 (1 件から 10 件)

10 件を表示 Previous Next

| データベース | データベース運用場所                            | 代表者       | データベースカテゴリ | 生物種 | 要約（キーワードを太字表示）   | 利用許諾          |
|--------|---------------------------------------|-----------|------------|-----|--|---------------|
|        | NBDC NikkajiRDF<br>ダウンロード<br>オリジナルサイト | 科学技術振興機構  | 化合物        | -   | 日本化学物質辞書（日化辞）のデータを化合物情報のRDF記述で標準となっているオントロジーを用いてRDF化したデータベース | CC表示<br>詳細    |
|        | RGP caps<br>ダウンロード   簡易検索<br>オリジナルサイト | 農業生物資源研究所 | DNAマーカー    | イネ  | イネの多型検出のために開発されたDNAマーカー（STSマーカーおよびCAPSマーカー）のデータベース           | CC表示-継承<br>詳細 |
|        | RGP estmap2001                        |           |            |     | YACデータのイネ転写産物地図化   |               |

102件のデータデータベース、約500件のデータセットを収録（2015年6月現在）。

# データベースとそのメタ情報

The screenshot shows the NCBI database entry for a genetic linkage map. It includes sections for 'データベースの概要' (Database Overview), 'データベースの解説' (Database Description), and 'データベースのリンク' (Database Links). The 'データベースの解説' section contains the title, authors (Y. Nagamura et al.), journal information (Genetics, 2001), and a list of keywords. The 'データベースのリンク' section lists various related databases like GenBank, EMBL, and UniProt.

# データセットとそのメタ情報

This grid displays several individual data set pages from the NCBI database. Each page follows a similar layout with sections for 'データ説明' (Data Description), 'データファイル' (Data Files), '検索方法' (Search Methods), and 'データ詳細' (Data Details). The pages shown include:
 

- 高品質連続地図情報**: A high-quality continuous map with 117 markers.
- 遺伝子型データ**: Genotype data for a specific marker, including a table with columns for '項目名' (Item Name) and '項目の説明' (Item Description).
- 連鎖順序情報**: Linkage order information for a set of markers.

# データセットのリスト

**ダウンロード**

はじめに利用許諾をお読みください。  
 なお、ここに表示したデータ名およびその説明は、生命科学データベースアーカイブから得たものです。オリジナルのデータベースの内容とは一致しない場合があります。

(「データ名」のリンクをクリックすると、データ項目の説明を参照できます。

| # | データ名           | データファイル                                 | 検索&ダウンロード |
|---|----------------|---|-----------|
| 1 | README         | README.html                             | -         |
| 2 | データのリンク先一覧     | rgg_gmap98_main.zip (355 B)             | 検索&ダウンロード |
| 3 | 高品質連続地図情報      | rgg_gmap98_detail.zip (37.2 KB)         | 検索&ダウンロード |
| 4 | アイソザイムの遺伝子型リスト | rgg_gmap98_isozyme_list.zip (611 B)     | 検索&ダウンロード |
| 5 | 遺伝子型データ        | rgg_gmap98_genotype_data.zip (48.4 KB)  | 検索&ダウンロード |
| 6 | 連鎖順序情報         | rgg_gmap98_linkage_order.zip (8.7 KB)   | 検索&ダウンロード |
| 7 | 連鎖順序画像         | rgg_gmap98_southern_image.zip (69.2 MB) | -         |

• FTPサイトからダウンロード

個々のデータセットは、説明のページ(固有のURL)を持ち、メタデータ、データセットへのリンク、利用許諾(クリエイティブ・コモンズ)などの情報を収録。

# 3. クリエイティブ・コモンズ (CC)

## 再利用と共有を促進する仕組み

### CCライセンスの種類

作品の利用（再配布やリミックス作品の公開、実演等）のための条件は4種類あります。



表示

作品のクレジットを表示すること



非営利

営利目的での利用をしないこと



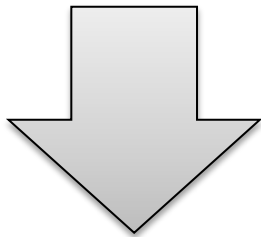
改変禁止

元の作品を改変しないこと



継承

元の作品と同じ組み合わせのCCライセンスで公開すること



4種類のマークを組み合わせると6種類の利用条件を選択することができる



CC-BY ジャパンクリエイティブ・コモンズ・ジャパン (<http://creativecommons.jp/licenses/>)

生命科学系データベースアーカイブで用いられる標準的なライセンス

# 4. テストの方法

- 登録対象データと登録対象件数
  - 生命科学系データベースアーカイブに収録されているデータセット全件（約500データセット（102データベース））
  - 登録対象件数（登録実験済の数）：74データセット（17データベース）
- テスト登録方法
  - 登録のインターフェースは、Web入力、XML一括入力の両方を使用した。
  - RAは、JaLCを選択。

# 実施体制

NBDC (バイオサイエンスデータベースセンター)  
「研究データへの DOI 登録実験プロジェクト」チーム

責任者：櫛田研究員 (NBDC研究員、ライフサイエンスDB統合のための基盤整備)

- ・ JaLCとの連絡
- ・ DOI登録の方式の検討
- ・ データ登録、更新、削除
- ・ 成果報告および報告書作成

生命科学系データベースアーカイブチーム (3名)

- ・ メタデータ作成



## 5. 結果 (1/2)

- 生命科学系データベースアーカイブのメタデータが、新規作成、修正することなく、ほぼそのままの形でJaLCの研究データ登録に流用できることを確認した (表)。

# 表. メタデータ項目の比較と値の一例

| メタデータ項目 (主なもの)   |                  | JaLCメタデータの値   |
|------------------|------------------|---|
| 生命科学系データベースアーカイブ | JaLCのメタデータ       |   |
| データ名             | title            | RGP gmap98 - 遺伝子型データ  |
| データURL           | url              | <a href="http://dbarchive.biosciencedbc.jp/jp/rgp-gmap98/data-4.html">http://dbarchive.biosciencedbc.jp/jp/rgp-gmap98/data-4.html</a> |
| 作成者氏名            | creator (ja)     | 春島嘉章、倉田のり、長村吉晃、矢野昌裕、佐々木卓治*  |
| 作成者英名            | creator (en)     | Yoshiaki Harushima, Nori Kurata, Yoshiaki Nagamura, Masahiro Yano, Takuji Sasaki  |
| 作成者所属            | affiliation      | 社団法人農林水産先端技術研究所、農業生物資源研究所、<br>社団法人農林水産先端技術研究所、農業生物資源研究所、<br>農業生物資源研究所   |
| 利用許諾             | rights_url       | <a href="http://dbarchive.biosciencedbc.jp/jp/rgp-gmap98/lic.html">http://dbarchive.biosciencedbc.jp/jp/rgp-gmap98/lic.html</a>       |
| 運用開始年月日          | publication_date | 1998/1/17   |
|                  | DOI              | 10.14977/01.rgp_gmap98_genotype_data  |
|                  | publisher        | 国立研究開発法人 科学技術振興機構<br>Japan Science and Technology Agency (JST)  |

## 5. 結果 (2/2)

- ランディングページは、生命科学系データベースアーカイブのデータセットの「説明ページ」が適当であることを確認した。

# ランディングページの例 (ダウンロード)

RGP gmap98

データベース名

このデータベースについて

- データベースの説明
- ダウンロード
- 利用許諾
- このデータベースの更新履歴

遺伝子型データ

データセット名

| データ説明        |  |
|--------------|--|
| データ名         | 遺伝子型データ  |
| データ内容の説明     | 各染色体の全ての分離個体 (186個体) についての遺伝子型データ  |
| データファイル      | データファイル名: rgp_gmap98_genotype_data.zip<br>データのURL: <a href="ftp://ftp.biosciencedbc.jp/archive/rgp-gmap98/LATEST/rgp_gmap98_genotype_data.zip">ftp://ftp.biosciencedbc.jp/archive/rgp-gmap98/LATEST/rgp_gmap98_genotype_data.zip</a><br>ファイルサイズ: 48.4 KB |
| 簡易検索URL      | <a href="http://togodb.biosciencedbc.jp/togodb/view/rgp_gmap98_genotype_data">http://togodb.biosciencedbc.jp/togodb/view/rgp_gmap98_genotype_data</a>  |
| データ取得方法      | NipponbareとKasalathの掛け合わせから得た186のF2植物体を用いて連鎖解析を行った結果。  |
| 解析方法         | MAPMAKER/EXP 3.0を使用して、連鎖解析を行いました。  |
| データ件数        | 2,277 件  |
| データ詳細        |  |
| 項目名          | 項目の説明  |
| Chrom. No.   | 染色体番号  |
| Locus        | 遺伝子座名 (マーカー名)  |
| Genetic code | MAPMAKERによる遺伝コード   |

DOI

[10.14977/01.rgp\\_gmap98\\_genotype\\_data](https://doi.org/10.14977/01.rgp_gmap98_genotype_data)

ランディングページのURL  
(データURL)

<http://dbarchive.biosciencedbc.jp/jp/rgp-gmap98/data-4.html>

# ランディングページの例 (簡易検索)

|         |   |
|---------|---|
| 簡易検索URL | http://togodb.biosciencedbc.jp/togodb/view/rgp_gmap98_genotype_data |
| データ取得方法 | NipponbareとKasalathの掛け合わせから得た186のF2植物体を用いて連鎖解析を行った結果。               |
| 解析方法    | MAPMAKER/EXP 3.0を使用して、連鎖解析を行いました。                                   |

English Japanese

RGP gmap98 - 遺伝子型データ NBDC - アーカイブトップ - ヘルプ

データベースの説明 | データ項目の説明 | 利用許諾 | ヘルプ

データのリンク先一覧 | 高密度連鎖地図情報 | アインザイムの遺伝子座リスト | [遺伝子型データ](#) | 連鎖順序情報

< 利用者の方へ >

- ダウンロード (Download:の右) する前に[利用許諾](#)を注意深くお読み下さい。ダウンロードを実行することによって、本利用許諾を承諾したものと見なします。

Find  All Search Clear [Advanced search](#)

Download: Selected as CSV [All](#)

| Entry                | Chrom. No. | Locus | Genetic code  |
|----------------------|------------|-------|---|
| <a href="#">Show</a> | 1          | R687  | HHBHAHHAAABVHAHHHHHHBHHBAHHHVBHAHVBHHHHAHVBHBA  |
| <a href="#">Show</a> | 1          | C161  | HHBHAHHAAABVHAHHHHHHBHHBAHHHVBHAHVBHHHHAHVBHBA<br>BHHABHH-BHHHAH  |
| <a href="#">Show</a> | 1          | C602  | H-BHAH---ABHAAHH-HHBHHBAH-<br>HBHAHVBHHHHAHVBHBAHHHVBHBAHHHVBHBAHHVBABBHAAHHHHI<br>HBABBHHHVBHAHABHBBAAHVBHAHAAHHHHHVBHHAHVBBAHVB   |
| <a href="#">Show</a> | 1          | S1442 | H-B-AH---ABHAAHH-HHBHAHAH-HBHAHVBHA-<br>HAHVBHBAHHHVBHAAHHHVBAAHHVBABBHAAAHH-<br>HHAHAAHAAHAAHAAHAAHAAHVBHHHHHH-<br>HBHVBHHHVBHAHABHBBAAHAAHBA--<br>AHVBHVBHBAHBBAAHVBHHAHVBHVBHVBHVBHVBHVBHVBHAA |
| <a href="#">Show</a> | 1          | C970  | HHBHAHHAAHABVHAHHHHHHBHAHAAHHHVBHAHVBHAAHHAHVBHBA   |

# ランディングページの例 (利用許諾)

RGP gmap98

## このデータベースについて

- データベースの説明
- ダウンロード
- 利用許諾**
- このデータベースの更新履歴

## 遺伝子型データ

### 利用許諾

#### データ説

#### 本データベースの利用許諾

利用許諾更新日: 2016/02/12

#### データ名

本データベースは、以下で定める利用許諾に基づきご利用いただくことができます。本利用許諾は、本データベース利用における許諾内容、及び利用者が従うべき条件を定めています。

#### データ内

#### データフ



本データベースの利用許諾は、クリエイティブ・コモンズ 表示-継承2.1 日本のも定める利用許諾です。本データベースのクレジットは、"RGP: A HIGH-DENSITY RICE GENETIC MAP © 農業生物資源研究所 licensed under CC表示-継承2.1 日本"ですので、利用にあたり必ず表示してください。

#### 簡易検索

クリエイティブ・コモンズ 表示-継承2.1 日本のご要は [こちら](#)です。具体的な許諾条項は [こちら](#)をご覧ください。

#### データ取

本データベースにおいて、以下の条件に従う限り許諾されている事項:

#### 解析方法

#### データ件

#### データ詳

1. 本データベースの全部または一部に自由にアクセスし、データを取得することができます。
2. 本データベースの全部または一部のデータを自由に再配布することができます。
3. 本データベースの全部または一部のデータを利用した、データベースなどの二次的著作物を自由に作成し、配布することができます。

本利用許諾に基づいて利用する際に従うべき条件:

Chrom. N

Locus

Genetic c

1. 本データベースの全部または一部、あるいは二次的著作物の配布に際しては、本データベースの作成者のクレジットを表示しなければなりません。
2. 本データベースの全部または一部のデータを利用して作成された二次的著作物は、この利用許諾の下で配布されなければなりません。
3. 本利用許諾で許諾されていない事項については、以下のデータベース作成者に連絡をとり、利用許諾を求める必要があります。

データベース作成者連絡先:

〒305-8602 茨城県つくば市観音台2-1-2

国立研究開発法人 農業生物資源研究所

松本隆

E-mail: [mat@affrc.go.jp](mailto:mat@affrc.go.jp)

片寄裕一

E-mail: [katayose@affrc.go.jp](mailto:katayose@affrc.go.jp)

型データ

p-  
genotype\_data.zip

genotype\_data

物体を用いて連鎖解析を行った結果。

## 6. 課題と検討事項

- Suffixのデザインをどうするか。人間が可読なもの（データセットの名称を含むSuffix）にするか否か（シリアル番号）について検討中。

10.14977/01.rgp gmap98 genotype data

Prefix

Suffix

- JaLCのメタデータ項目の作成者および寄与者のそれぞれにどの情報を入力すべきか結論が出ていない。
- データ更新時の新DOIの付与や新旧データの保管が必要になった場合へ対応が決まっていない。

# 7. 本番登録の予定

- 前述の課題を解決した時点で、登録を開始したい。