

### 統合データベース講習会: AJACS本郷10

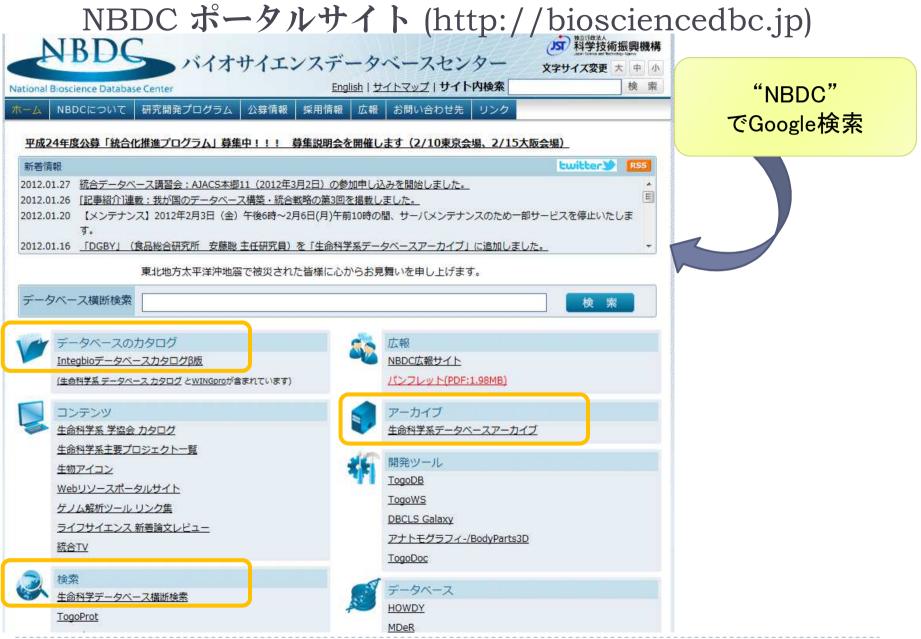
### NBDCの三大サービス (カタログ、横断検索、アーカイブ)の紹介

2012年2月10日(金)

バイオサイエンスデータベースセンター 櫛田達矢

### テーマ

- ▶ NBDC 三大サービス
  - ▶ Integbioデータベース<u>カタログ</u>
  - ▶ 生命科学データベース<u>横断検索</u>
  - ▶ 生命科学系データベース<u>アーカイブ</u>
- ▶ NBDC提供 その他のサービス



# Integbioデータベース<u>カタログ</u>

## Q. Integbioデータベース<u>カタログ</u>とは?

### A. 国内外の生命科学系データベースのリスト(カタログ)

- 対象のデータベース:1088件
  - ▶ 国内のデータベース:600件以上
- ▶ データベースのメタ情報を収集、整理
  - ▶ 所在情報(名称、URL、国など)、言語、内容説明、生物種、更新日、など
- ▶ データベースの情報を閲覧、検索、ダウンロードすることが可能
- ▶ 各機関のデータベース情報を集約
  - ▶ NBDC:「生命科学系データベースカタログ」、「WINGpro」
  - ▶ MEDALS:「データベース便覧」
  - ▶ AgriTOGO:「データベース便覧」
  - ▶ 医薬基盤研究所:「各種データベース」

### NBDCポータルサイト

## → Integbioデータベース<u>カタログ</u>



## Integbioデータベースカタログ 「探す」1/2

#### stintegbio.jp 「DBカタログ事業」へ Integbio データベースカタログ β版 1.0 ~ データベースを探す 最終更新日:2012/02/06 NBDC「生命科学系データベース力タログ」、「WINGpro」、MEDALS「データベース便覧」、AgriTOGO「データベース便覧」、医薬基盤研究所「各種データベース」の情報を集約しています。 国内外の生命科学系データベースの情報を閲覧、検索、ダウンロードすることが可能です <ご利用にあたって>【Show】レコードの詳細表示【Show Advanced Search】詳細検索【一覧 能です。 【本力タログについて】 【利用許諾とお問い合せ】 Displayed columns <新着情報> 2011/12/12 Integbioデータベースカタログ 8版1.0を公開しました。 Show Advanced Search 発現 ▼データベース名称とURL Columns ... 15 🔻 123.73 Next □ データベースID 1088 Found - データベース名称 データベー ス名称 とURL 製明 運用機関名 8 略称 棋 URL ☑ 稼働状況 (2/6) 国内外の生 ベースについて 言語 の説明や4 ています。 Integbioデータベースカタログ 2011年12 運用機関ID - スカタログの 独立行政法人科学技術 稼働中 [略称:] エントリー Show ☑ 運用機関名 振興機構 http://togodb.dbcls.jp/integbio dbcatalog •文部科学 運用機関所在国 [WINGpr ドータベース便 覧」。厚生的 ク集 生物種 V ≣#8A 生体分子 タベース (PDB)を中 生命科学系データベースカタログへのリンク 3DinSight SWISSP に統合しまし [略称:] 九州工業大学 稼働中 ■ WINGproへのリンク Show た。構造、 則べることがで http://gibk26.bio.kyutech.ac.jp/jouhou/3dinsight/3DinSight.html ■ 生命科学系データベースアーカイブへのリンク きます。構 機能部位をマ ッピングし ■ MEDALSへのリンク □カタログデータ最終更新日時 又録したデータ **廿細胞株** ベースです OK Cancel 5'-end serial analysis of gene expression database オリジナル bpの断片を用 東京大学 大学院新領 FOR Dr - C'CACET

## Integbioデータベースカタログ 「探す」2/2

#### Integbio データベースカタログ β版 1.0 ~ データベースを探す

置integbio.jp「DBカタログ事業」へ

最終更新日:2012/02/06

ギータベーフID	データベース 名称	TAL AND	URL	稼働状況(	<del>=</del> ===================================	運用機関ID	a 田 桦 関 夕	運用機関所在国	生物種	説明
7 -3 \ - \ 10	) - 3 \ - A -a 10	<b>ሥ</b> ር የጥ	OKL	119K 198/0 17/17/17/11 (	<b>=</b> •••	建州版は近	建用 版因	建用版制加红色	工作列生	0元971
NBDC00003	5'-end serial ana	5'SAGE	http://5sag	稼働中	en	20090509837	東京大学 オ	Japan	Homo sapie	ヒト細胞株
NBDC00006	Archaeal Gene N	Arch GeNe	http://riodl	稼働中	en	2009050472 26407099/A 142000000	独立行政法。	Japan	Thermoplas	好熱性古紅
NBDC00009	ArrayExpress Arc	ArrayExpre:	http://wwv	稼働中	en		European Bio	United Kingdom		遺伝子発現
NBDC00014	Berkeley Drosop	BDGP	http://wwv	稼働中	en		The Berkele	United States	Drosophila	ショウジョウ
NBDC00020	Bacillus subtilis (	BSORF	http://baci	稼働中	en	20090507435	京都大学 化	Japan	Bacillus su	JAFAN(枯
NBDC00025	Cancer Gene Ex	CGED	http://lifes	稼働中	en	20090509676	情報・システ	Japan	Homo sapie	癌の遺伝子
NBDC00029	Center for Inforr	CIBEX	http://cibe	稼働中	en	2009050641 23833172/A 158004090	情報・システ	Japan		CIBEXIX、
NBDC00038	Database of the	RNAi phend	http://park	稼働中	en	20090503832	東京大学	Japan	Caenorhab	生殖細胞(
NBDC00042	SNPs in the tran		http://gsnp		en	20090502617		Japan	Homo sapie	
NBDC00044	Database of Tun	DBTGR	http://dbtg		en	20090504111			Ciona/771	
NBDC00055	ENTREZ, Life Sc	iences Sea			en			United States		NCBI Entre
NBDC00065	A Drosophila Ima		http://flyvi		en		University of	Germany	Drosophila	ショウジョウ
ttp://www.fruitfly.		14丁光初	ハス・ノチのチー	xezimotoay	· ンローンヤフ۹ン	ンツの効能でいっている。	9 .	Project	1	ton Ethio

# 生命科学データベース横断検索

## Q. 生命科学データベース横断検索とは?

### A. 生命科学分野を対象にしたGoogleライクな検索

← 情報源を生命科学に限定。正解率の高いデータを提供



Googleで「ips」を検索した場合



網羅的だが、意図する回答と異な る結果も多い

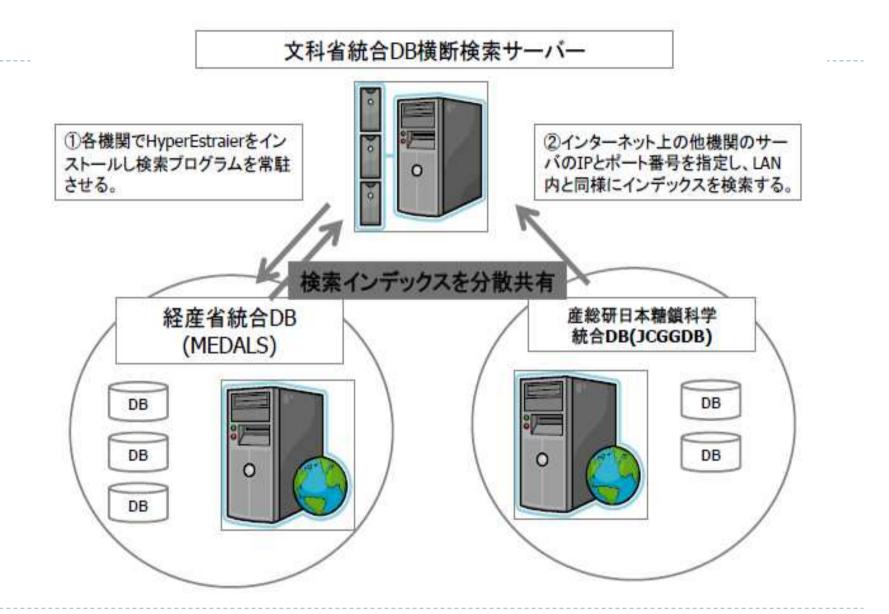
## 生命科学データベース横断検索の特徴 1/2

- 対象のデータベース:300件以上
- ▶ 分子DBだけでなく、雑誌(蛋白質核酸酵素(PNE))や特 許公開公報なども検索対象
  - ▶「蛋白質核酸酵素」のPDFは共立出版株式会社からの協力と 提供
- ▶ 検索キーワードの日英相互翻訳
  - ライフサイエンス辞書(京都大学)
- ▶ DBカテゴリを利用した絞り込み機能

## 生命科学データベース横断検索の特徴 2/2

- ▶ 文科省、経産省のDBに対して、共通の横断検索システムを採用。統一的な横断検索を実現。
  - ▶ 厚労省、農水省にも拡大する予定
  - ▶ NCBI(中央集権型システム)とは異なる
- 検索エンジンとして、全文検索システム「Hyper Estraier」 を使用。
  - http://fallabs.com/hyperestraier/

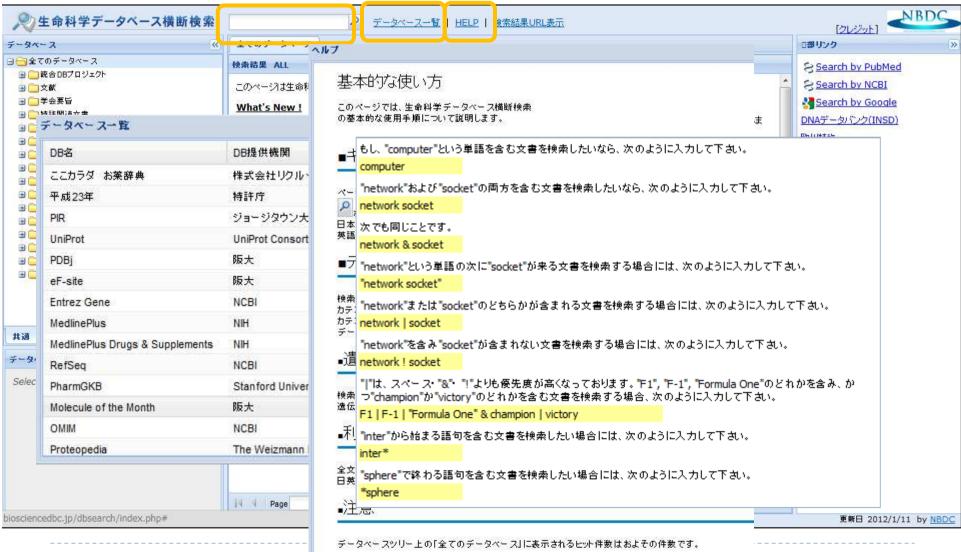
## 横断検索の連携システムの概要



## NBDCポータルサイト →生命科学データベース<u>横断検索</u>



## 生命科学データベース横断検索「検索する」1/4



## 生命科学データベース横断検索「検索する」2/4

### 検索語の入力方法

- 「network」および「socket」の両方を含む文書を検索する場合
  - network socket
  - network & socket
- 「network」という単語の次に「socket」が来る文書を検索する場合
  - "network socket"
- 「network」または「socket」のどちらかが含まれる文書を検索する場合
  - network | socket
- 「network」を含み「socket」が含まれない文書を検索する場合
  - network! Socket
- ▶ 「inter」から始まる語句を含む文書を検索したい場合。
  - inter\*
- ▶ 「sphere」で終わる語句を含む文書を検索したい場合。
  - \*sphere

## 生命科学データベース<u>横断検索</u>「検索する」3/4



DBの詳細情報

## 生命科学データベース<u>横断検索</u>「検索する」4/4

### > 実習1

▶ 検索キーワード「新型インフルエンザ」で検索

### > 実習2

▶ 検索キーワード「COX2」で検索

### > 実習3

▶ 検索キーワード「ヒアルロン酸」で検索

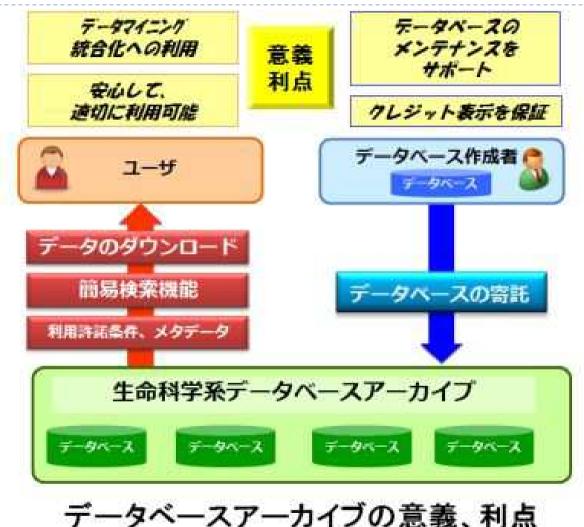
# 生命科学系データベース<u>アーカイブ</u>

### Q. 生命科学系データベース<u>アーカイブ</u>とは? 1/2

A. データベースを丸ごとダウンロードできるサービス ← データの最大限の再利用が可能に。

- データベース数:51件
- ▶ ダウンロードされるデータは表形式(Excelで読み込み可能)
- クリエイティブ・コモンズ・ライセンスに基づく利用許諾が与えられる。

## Q. 生命科学系データベース<u>アーカイブ</u>とは? 2/2



データベースアーカイブの意義、利点

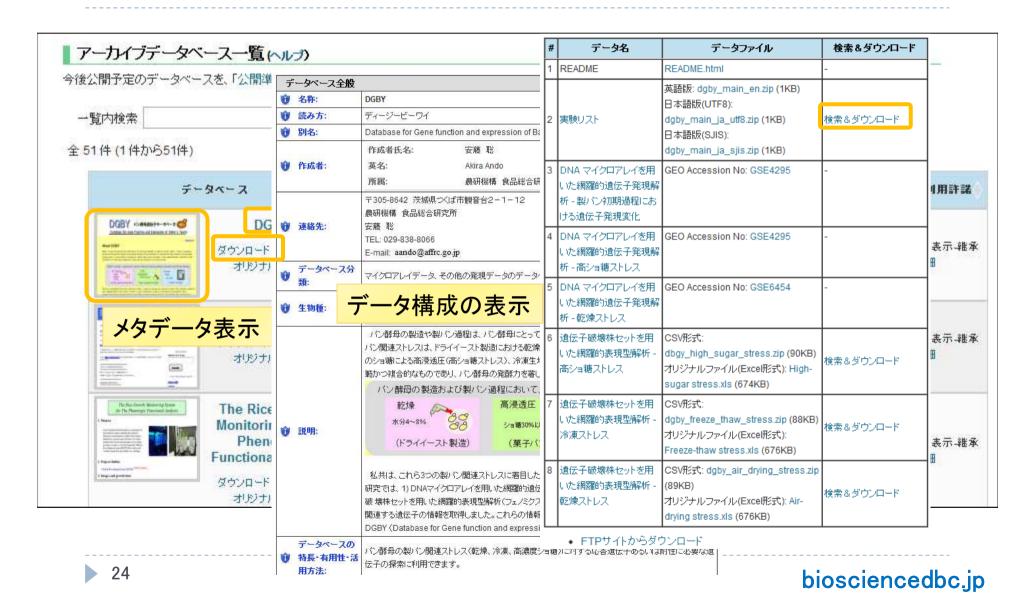
## NBDCポータルサイト →生命科学系データベースアーカイブ



### 生命科学系データベースアーカイブ「ダウンロードする」1/4

Life Science Date	abase Archi			-あのデータベー				アーカイブ横断検索	
Arc				生命科学系	データベ	ースアーカ	カイブ	(+a.±)	
								[検索]	
ホーム・アール	カイブの裏	(明 寄	化店套要領	更新風歴 ^	いけ 調	い合わせ			
いくら良質なデータ い、ダウンロードで い評価をうける機会	きないなと	の理由で十	分に利用され	用条件が明確でな 1、引用され、相応し	データフィニングM 安心して、通	OCPURAL	タベースの保持者項 クレジット相示	新アーカイブ情報 2012/01/16 [DGBY] (	(食品総合研究
生命科学系デー 者が生み出したデ に維持保管し、デ	<b>タベース</b> ータセット ータ説明()	<b>アーカイブ</b> は をわが国の公 ×タデータ)を	、国内のラ- 注財として 統一して検索	まとめて長期間安定	F-909		タベース作成者	所 安藤聡 主任研究員 た 2011/12/02 [Fluorom	
アクセスレダウンロ	コードを行え	とるようにする	サービスで	す(詳細説明)。	<b>総規模</b> 中本記代IUH			園池公穀教授)のデータ 更新しました	文をver 2.01
- At E 101-1-	(-KUZ)				The second secon	and the latest terminal termin	-	更新しました	
	や民間等権	葉々なユーザ	が利用しや	すい形にすること	9:	科学系データベース	アーカイブ	2011/11/17 [INOH] (1	the INOH
る一方、公的機関	や民間等権 究の生命科	葉々なユーザ 4学へのいっ	が利用しや そうの貢献を	すい形にすること 支援します。 デー	<b>1</b>	<b>科学系データベース</b>	アーカイブ	2011/11/17 [INOH] (i database project) を追	
る一方、公的機関 で、それぞれの研?	や民間等権 究の生命科 随時募集し	様々なユーザ 斗学へのいっ ンています(寄	が利用しや そうの貢献を 託応募要領	すい形にすること 支援します。 デー	<b>1</b>	8科学系データベース	アーカイブ	Serve Marry Lawrence and Control of the Control of	
るー方、公的機関で、それぞれの研究 タベースの <b>容</b> 託を ア <b>ーカイブデ</b>	や民間等権 究の生命権 随時募集し	<sup>集々なユーザ</sup> 科学へのいっ これ、ます(者 <b>一人一覧</b>	が利用しや そうの貢献を 託応募要領 (ヘルプ)	すい形にすること 支援します。 デー		日科学系データベース	アーカイブ	database project) を追	加しました
る一方、公的機関 で、それぞれの研究 タベースの <b>寄</b> 託を ア <b>ーカイブデ</b>	や民間等権 究の生命権 随時募集し	<sup>集々なユーザ</sup> 科学へのいっ これ、ます(者 <b>一人一覧</b>	が利用しや そうの貢献を 託応募要領 (ヘルプ)	すい形にすること  支援します。 デー  )。		科学系データベース	アーカイブ	Serve Marry Lawrence and Control of the Control of	加しました
る一方、公的機関 で、それぞれの研究 タベースの <b>寄</b> 託を ア <b>ーカイブデ</b> 今後公開予定のデ	や民間等を 究の生命を 随時募集し <b>データベ</b>	<sup>集々なユーザ</sup> 科学へのいっ これ、ます(者 <b>一人一覧</b>	が利用しや そうの貢献を 託応募要領 (ヘルプ)	すい形にすること  支援します。 デー  )。		科学系データベース	アーカイブ	database project) を追	加しました
る一方、公的機関で、それぞれの研究タベースの寄託を図 タベースの寄託を図 アーカイブデ 今後公開予定のデーー覧内検索	や民間等解 で で で で で で で で で で で で で	<sup>集々なユーザ</sup> 科学へのいっ これ、ます(者 <b>一人一覧</b>	が利用しや そうの貢献を 託応募要領 (ヘルプ)	すい形にすること  支援します。 デー  )。	ಸು≢ತ.			database project) を追	加しました
る一方、公的機関で、それぞれの研究タベースの寄託を図 タベースの寄託を図 アーカイブデ 今後公開予定のデーー覧内検索	や R B B B B B B B B B B B B B B B B B B	様々なユーザ 科学へのいっ たれ、ます(寄 一 <b>ス一覧</b> を、「公開準	が利用しや そうの貢献を 託応募要領 (ヘルプ) 備中のデータ ・ ・ ・ ・ ・ ・ ・ ・ ・ ・ ・ ・ ・ ・ ・ ・ ・ ・ ・	すい形にすること :支援します。 デー )。 タベース一覧 JI 2掲載して	ています。 <b>場所 代表</b> を 食品産	データベー: カテゴリ		database project) を追 並び替えが	可能利用許認

### 生命科学系データベース<u>アーカイブ</u>「ダウンロードする」2/4



### 生命科学系データベースアーカイブ「ダウンロードする」3/4

Die	BY - §	利用許諾	ルブ
N. Parketo		■本データベースの利用許諾	
	データベ・	利用許諾更新日: 2012/01/16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。	
• 5	利用者の: 'ウンロード ow Advan	<b>② ② ③</b> 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 日本の定める利用許諾です。 本データベースのクレジットは、 <b>"パン酵母遺伝子データベース ◎ 食品総合研究所 licensed under CC表示 継承2.1 日本"</b> ですので、利用にあたり必ず表示してください。	Columns
ID	ポストゲン ローチ	クリエイティブ・コモンズ 表示・継承2.1 日本の概要はこちらです。具体的な許諾条項はこちらをご覧ください。	
1	網羅的造(	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:	Show
2	網羅的造作析	<ol> <li>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。</li> <li>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。</li> </ol>	Show
3	網羅的遗作	<ol> <li>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。</li> </ol>	Show
4	網羅的進作	本利用許諾に基づいて利用する際に従うべき条件:	Show
11	網羅的表現	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	Show
12	網羅的表現	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。	Show
13	網羅的表現	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。	Show

### 生命科学系データベースアーカイブ「ダウンロードする」4/4

#### DGBY - 実験リスト

NBDC - アーカイブトップ - ヘルブ

データベースの説明 | データ項目の説明 | ダウンロード | 利用許諾 | ヘルブ

#### 実験リスト: 日本語版 | 英語版

- 4	Α	В	С	D	E	F	G	Н	I	J	K
1	ORF	GENE	Growth rate YM/YW	Growth rate SM/SW	DESCRIPTI	ON (MIPS)					
2	YALO02W	VPS8	0.99	0.91	vacuolar sorting protein, 134 kD						
3	YALO04W		1.05	1.01	questionable protein						
4	YAL005C	SSA1	1.05	1.11	heat shock	protein of	HSP70 fan	nily, cytosoli	ic		
5	YAL007C	ERP2	1.12	1.01	p24 protein involved in membrane trafficking						
6	YALO08W	FUN14	1.1	0.97	Protein of	unknown fu	unction				
7	YALO09W	SPO7	1.04	0.75	meiotic pro	tein					
8		MDM10	0.33		involved in						
9	YALO11W		1.01		component			that incor	porates Htz	1 p into chr	omatin
10	YAL012W		0.97		cystathioni						
11		DEP1	0.87		regulator of			sm			
12		SYN8	1.01		endosomal		tein				
13		NTG1	0.97		DNA repair						
14	YALO16W		0.73		ser/thr pro						
15		PSK1	1.04		PAS kinase						tion
16	YALO18C		1.02		weak simila						
17	YALO19W		1.06		protein imp			e integrity a	and segregat	ion	
18		ATS1	0.99		alpha-tubul						
19					6 transcriptional regulator						
20	YAL022C				Broad-specificity nucleoside transporter of i		nternal men	nbranes			
21	YAL023C	PMT2	1.01		mannosyltm						
22		LTE1	ND	ND	GDP/GTP	_					/ eu
23			0.92		P-type ATF			•		translocas	a (flippase
24	YAL027W		0.98	1.0/	Protein of	unknown fi	unction loca	alised to nu	cleus		
	YAL017\	W PSK1	1.04 1.	12 PAS	kinase involv	ved in the c	ontrol of su	gar metabol	ism and trar	nslation	Show
	YAL0180	c -	1.02 1.0	01 weak	similarity to	hypothetic	al proteins	YOL047c an	d YOL048c		Show

# NBDC提供 その他のサービス

### NBDC提供サービス: <u>ゲノム解析ツール リンク集</u> (1/2)



#### データベースのカタログ

Integbioデータベースカタログβ版

(生命科学系 データベース カタログ とWINGproが含まれています)



#### コンテンツ

生命科学系 学協会 カタログ

生命科学系主要プロジェクト一覧

生物アイコン

Webリソースポータルサイト

ゲノム解析ツール リンク集

ライフサイエンス 新着 分子生物学に関わるデータ解析ツール提供サイトへのリ ンクと簡単な解説を提供するリンク集です



統合TV

生命科学データベース横断検索

TogoProt

GenLibi

Allie

inMeXes

SRAs

鎖鋸

統合遺伝子検索 GGRNA



#### 広報

NBDC広報サイト

パンフレット(PDF:1.98MB)



#### アーカイブ

生命科学系データベースアーカイブ



#### 開発ツール

TogoDB

TogoWS

**DBCLS Galaxy** 

アナトモグラフィー/BodyParts3D

TogoDoc



#### データベース

HOWDY

**MDeR** 

RefEx



#### 連携

integbio.jp

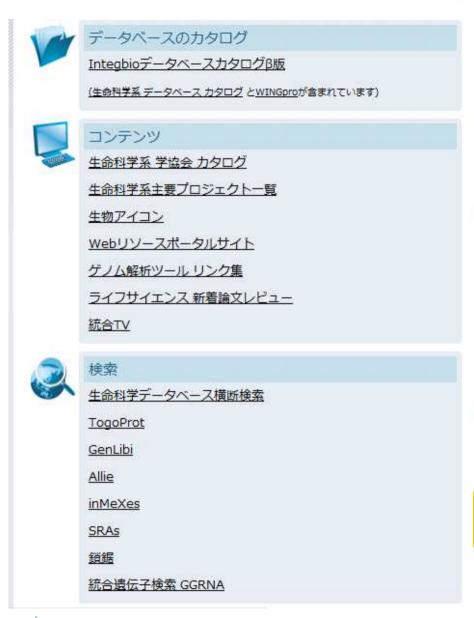
BioHackathon 2011、2010、2009、2008

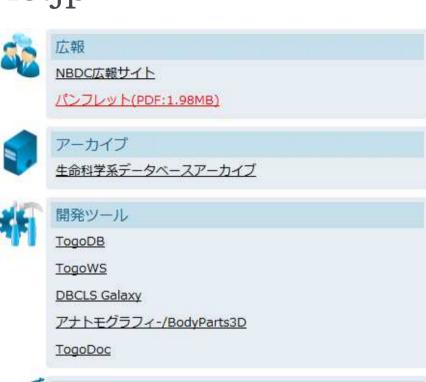
本ポータルサイトで提供しているサービスについて

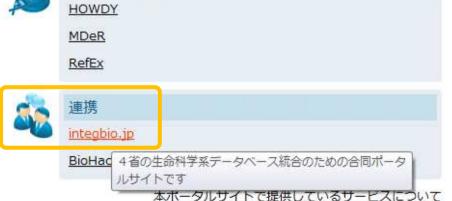
### NBDC提供サービス: <u>ゲノム解析ツール リンク集</u> (2/2)



## DB統合化の象徴: Integbio.jp

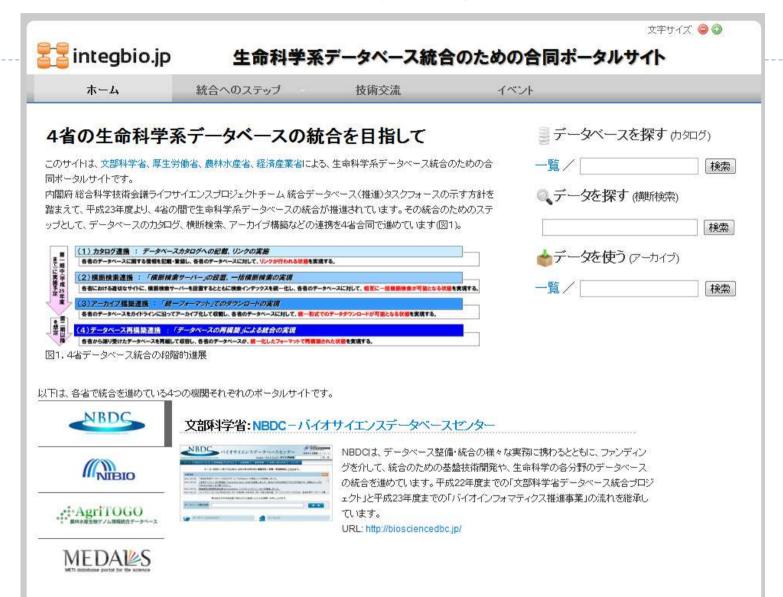






データベース

## DB統合化の象徴:Integbio.jp



### NBDC提供サービス:生命科学系 学協会カタログ(1/2)



### NBDC提供サービス:生命科学系 学協会カタログ(2/2)



## NBDC提供サービス: <u>生物アイコン</u>(1/2)



#### データベースのカタログ

IntegbioデータベースカタログB版

(生命科学系 データベース カタログ とWINGproが含まれています)



#### コンテンツ

生命科学系 学協会 カタログ

生命科学系主要プロジェクト一覧

#### 生物アイコン

Wentry- AN-Allina N

細菌から動植物まで200種類以上の生物種のアイコン

ゲノム解析 (イラスト) を無料でダウンロード・利用できます

ライフサイエンス 新着論文レビュー

統合TV



#### 検索

生命科学データベース横断検索

TogoProt

GenLibi

Allie

inMeXes

SRAs

鎖鋸

統合遺伝子検索 GGRNA



#### 広報

NBDC広報サイト

パンフレット(PDF:1.98MB)



#### アーカイブ

生命科学系データベースアーカイブ



#### 開発ツール

TogoDB

**TogoWS** 

**DBCLS Galaxy** 

アナトモグラフィー/BodyParts3D

TogoDoc



#### データベース

HOWDY

**MDeR** 

RefEx



#### 連携

integbio.jp

BioHackathon 2011、2010、2009、2008

本ポータルサイトで提供しているサービスについて

# NBDC提供サービス: <u>生物アイコン</u>(2/2)



## 参考資料

- 過去の講習会資料(三大サービス関係)
  - http://motdb.dbcls.jp/?AJACS27 (2011/7/20)
  - http://motdb.dbcls.jp/?AJACS21%2Fbono1 (2010/8/4)
  - http://motdb.dbcls.jp/?AJACS20%2Fshoko (2010/7/6)
  - http://motdb.dbcls.jp/?AJACS14%2Fsk2 (2009/11/6)
- ▶ 全文検索システム「Hyper Estraier」
  - http://fallabs.com/hyperestraier/
- ▶ 統合データベースプロジェクト(平成18年度~22年度)
  - http://lifesciencedb.mext.go.jp/

### Credit

- 以下のサービスは、文部科学省委託研究開発事業「統合データベース プロジェクト」の一環として、ライフサイエンス統合データベースセンター (DBCLS)で開発されたものを平成23年(2011)4月1日より、独立行政法 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が提供し ています。
  - 生命科学系データベースカタログ\*
  - 生命科学系 学協会カタログ
  - 生物アイコン
  - 生命科学データベース横断検索
  - 生命科学系データベースアーカイブ

\*:Integbioデータベースカタログは、NBDC「生命科学系データベースカタログ」、「WINGpro」、MEDALS「データベース便覧」、AgriTOGO「データベース便覧」、医薬基盤研究所「各種データベース」の情報を集約しています。