

- Site Map
- ご利用の前に
- DDBJとは?
- 塩基配列の登録
- SAKURA 大量登録
登録データ更新
- 検索・解析
- ARSA getentry
TXSearch
FASTA BLAST
SSEARCH HMMPFAM
ClustalW
Web API
GIB GIB-V
GTPS GTPOT
- 統計の詳細
- リリース取得
- DDBJリリースノート
公開中リリース
- Q and A
- 研究会
講習会 関連会議
ゲノム機能発現研究会
- DDBJの連絡先
www-admin@ddbj.nig.ac.jp
- Copyright © 1995-2006
DDBJ All rights reserved.

■暖かいですね

隔月公開の DDBJ メールマガジン第30号 web 版です。メールマガジンに関するご質問やご意見がありましたら ddbjmag@ddbj.nig.ac.jp までどうぞ。

■公開サービスの URL 変更について

国立遺伝学研究所では、本年3月にスーパーコンピュータシステムの新システムへの更新 (リプレイス) をおこないました。新システムへの移行に伴い、以下のサービスの URL を変更いたします。お手数ですが、リンク等の新 URL への更新をお願いいたします。当分の間、新旧いずれの URL でもアクセス可能 (検索解析サービスを除く) ですが、一定期間経過後、旧 URL は停止されます。予めご了承ください。詳細は、[「2007年2-3月におこなわれるサービスの中断・変更などについて」](#)をご覧ください。新システムへの移行により、DDBJ では、検索・解析サービスの高速化、新サービスの追加提供を実現いたします。ご迷惑とご不便をおかけいたしますが、ご理解とご協力をお願いいたします。

現在予定されているサービス中断等の日程

- 2月22日 (木) 17時から 2月28日 (水) 9時まで
[塩基配列データ登録受付・公開業務の停止について](#)
(SAKURA, TXSearch の一時停止)

URL が変更になるサービス

サービス名	旧 URL	新 URL
	TXSearch	http://sakura.ddbj.nig.ac.jp/uniTax-jp.html http://txsearch.ddbj.nig.ac.jp/
PDB Retriever	http://www.ddbj.nig.ac.jp/search/ http://pdbret.ddbj.nig.ac.jp/	
SSThread	2007/2/23 サービス終了	
CAMUS	2007/2/23 サービス終了	
Lib score	http://hypernig.nig.ac.jp/cgi-bin/Lib-score/ http://libscore.ddbj.nig.ac.jp/	
検索解析 トップ	http://www.ddbj.nig.ac.jp/search/index-j.html http://www.ddbj.nig.ac.jp/search/	
FASTA	http://www.ddbj.nig.ac.jp/search/fasta-j.html http://fasta.ddbj.nig.ac.jp/	
BLAST	http://www.ddbj.nig.ac.jp/search/blast-j.html http://blast.ddbj.nig.ac.jp/	
PSI-BLAST	http://www.ddbj.nig.ac.jp/search/psi_blast-j.html http://psiblast.ddbj.nig.ac.jp/	
SSEARCH	http://www.ddbj.nig.ac.jp/search/ssearch-j.html http://ssearch.ddbj.nig.ac.jp/	
ClustalW	http://www.ddbj.nig.ac.jp/search/clustalw-j.html http://clustalw.ddbj.nig.ac.jp/	
hmmpfam	新規サービス開始 http://hmmpfam.ddbj.nig.ac.jp/	
LIBRA	http://www.ddbj.nig.ac.jp/search/ http://libra.ddbj.nig.ac.jp/	

2月下旬に新システムへの切替を予定しておりますが、それに伴い、切替後の2月下旬から3月上旬位までの期間、DDBJ の HP を始め各種公開サービスのアクセスがうまくいかない場合があります。その場合には、時間を置いて再度アクセスしてください。ご迷惑をおかけいたしますが、ご理解とご協力をお願いいたします。

■DDBJ の新規データベース GIB-V と GTPS 公開

DDBJ では2つの新規データベース GIB-V (ウイルスゲノムデータベース) と GTPS (共通プロトコルに基づくバクテリアゲノムの再アノテーション) を公開しました。

▶ Genome Information Broker for Viruses (GIB-V) Genome Information Broker for Viruses (GIB-V) は、微生物ゲノムデータベース [Genome Information Broker \(GIB\)](#) の新しいシリーズとして構築されたウイルスゲノム/セグメントのデータベースです。DDBJ, EMBL および GenBank で構成される国際塩基配列データベース共同体 (INSDC) のデータベースより抽出した約25,000件の完全ウイルスゲノム/セグメントのデータを格納しています。

INSDC のデータベースにおいて、一部のゲノムプロジェクトからの登録データには、アクセッション番号の先頭に特定のプレフィックス (AE, AL, AP, BS, BX, CP, CR, CT, CU, CY) が付与され、ゲノムデータであることを判別できます。しかし多くのゲノムデータには通常の遺伝子と同じプレフィックスが割り振られているため、データベースからの抽出が難しくなっていました。

このため、ウイルス、ウィロイド、ファージの完全ゲノム/セグメントデータを極力網羅的に格納し提供することを目的として GIB-V を構築しました。

登録したウイルスのリストは分類の階層にしたがってまとめられており、クリックにより各階層を展開して表示していくことができます。各ゲノムのページではゲノム/セグメントデータが、グラフィカル表示やリスト、テキスト表示によって提示されます。

検索機能として、ウイルス名やアクセッション番号などで登録ウイルスを検索する Search for Virus information と、ORF 名やフィーチャー/クオリファイア情報などから ORF を検索する Search for ORF information の2種類を備えました。また BLAST による相同性検索機能も搭載しており、結果をグラフィックやリストの形で表示します。結果をオンラインで直接 DDBJ の ClustalW 解析にかけることも可能です。格納されているゲノム/セグメントデータを手元のコンピュータでダウンロードすることができますので、ご自分のコンピュータの内部で高度な解析をすることも可能です。

なお、データベース運用開発研究室では比較ゲノムツール G-InforBio (<http://wdcm.nig.ac.jp/inforbio/G-InforBio/download.html>) を開発しており、こちらとの連携も強化していく予定です。

GIB-V は <http://gib-v.genes.nig.ac.jp/> より無料でアクセスできます。登録は不要です。

(GIB-DDBJ データベース運用開発研究室 平畠壮規)

Gene Trek in Prokaryote Space (GTPS)

GTPS では、統一されたアノテーションプロトコルに基づいて、公開されている真正細菌および古細菌由来のゲノムに対して一斉に再アノテーションを行い、そこで予測された全タンパク質コード領域 (ORF) のデータを提供しています。

既存のゲノムデータにおいては、ORF 予測のために種々の設定 (最低 ORF 長さや相同性検索における設定など) が用いられています。また、参照データベースのバージョンも新旧混用されています。このため、基準的な条件設定に基づく ORF 予測が行われておりません。その結果、ORF が見落とされている事例も生じています。GTPS で用いているアノテーションプロトコルでは、最低 ORF 長を 48 bp (15 aa) に設定した上で候補となる ORF をゲノム中から可能な限り探索し、共通の条件設定による相同性検索およびモチーフ解析を行っています。予測された ORF には、相同性解析とモチーフ検索の結果に基づく、信頼度に応じたグレードが与えられています。GTPS が提供する高グレードの ORF の中には、既存のゲノムデータには存在しないものが含まれる場合や、既存ゲノムデータに存在する ORF が、GTPS では遺伝子の可能性のない、低グレードとして判定されているものがあります。

GTPS では 2003年から全バクテリアゲノムへの再アノテーションを開始しており、以降、毎年更新を繰り返してきました。結果は GTPS データベース (<http://gtps.ddbj.nig.ac.jp/>) から取得することができます。GTPS データベースでは、グラフィカルビューソフトウェアを利用して ORF データの内容を閲覧することができ、各 ORF の詳細画面においては、相同性やモチーフ検索結果、および、予測 ORF どちらのオルゴリズム解析の内容を確認することができます。すべての結果はフラットファイル形式でも提供しております。

(DDBJ データバンク構築局 小菅武英)

■DDBJ の活動に関する論文の掲載

▶ Nucleic Acids Research (DDBJ の活動全般, GTPS, GIB-V に関して)

"DDBJ working on evaluation and classification of bacterial genes in INSDC"
Hideaki Sugawara, Takashi Abe, Takashi Gojobori and Yoshio Tateno
[Nucleic Acids Research, 2007, Vol. 35, Database issue D13-D15](#)

(内容) 2005年7月から2006年6月の間に DDBJ が登録を受け付け、公開したミツパチの ESTs, 病原性大腸菌 K-12 W3110, メダカ WGS, ヒト MGA のゲノムデータ等について。さらに、INSDC に登録されたバクテリア遺伝子の分析評価についても詳細に記載されています。
(PMID: 17108353)

▶ OMICS 2006 (TPA に関して)

"Evidence Standards in Experimental and Inferential INSDC Third Party Annotation Data"
Guy Cochrane, Kirsty Bates, Rolf Apweiler, Yoshio Tateno, Jun Mashima, Takehide Kosuge, Ilene Karsch Mizrach, Susan Schaffer, Michael Fetchko
[OMICS 2006, Vol. 10, No. 2: 105-113](#)

(内容) 国際塩基配列データベースを構築する日・英・米の3データバンクが、増加する TPA (Third Party Annotation) 登録についてのケーススタディをもとにした分析を共同執筆しました。

▶ DNA Research (GTPS について)

"Exploration and Grading of Possible Genes from 183 Bacterial Strains by a Common Protocol to Identification of NewGenes: Gene Trek in Prokaryote Space (GTPS)"

Takehide Kosuge, Takashi Abe, Toshihisa Okido, Naoto Tanaka, Masaki Hirahata, Yutaka Maruyama, Jun Mashima, Aki Tomiki, Motoyoshi Kurokawa, Ryutaro Himeno, Satoshi Fukuchi, Satoru Miyazaki, Takashi Gojobori, Yoshio Tateno, and Hideaki Sugawara
[DNA Research Advance Access published online on December 13, 2006.](#)

(内容) 上述

■ClustalW DOTSIOUTPUT オプションご利用のお知らせ

DDBJ が WWW と E-mail で提供している ClustalW のアライメント結果をドットで出力する (以下出力例) DOTSIOUTPUT オプションが利用可能になりました。DOTSIOUTPUT オプションは、ALIGN の解析のみ利用可能で、DOTSIOUTPUT オプションを ON にした場合は TREE や BOOTSTRAP 解析は実行できません。

```

CLUSTAL W (1.83) multiple sequence alignment
Query          MDMEIEDSSPIDDLKIQLDLVNYPQPCLELTQPILLQYENIKFIKIPVGNLSTEK
F39728-1       .....M.Q.QP.Q.SA...F.LNT.S.AEPIESK.R.F.SIGIPIQ.VVE
Q759A6-1       .....M.Q.QP.Q.SA...F.LNT.S.AEPIESK.R.F.SIGIPIQ.VVE
Q6FV15-1       .....M.Q.QP.Q.SA...F.LNT.S.AEPIESK.R.F.SIGIPIQ.VVE
Q12182-1       .....M.Q.QP.Q.SA...F.LNT.S.AEPIESK.R.F.SIGIPIQ.VVE
Q6CIX1-1       .....LNQM.P.Q.SN...F.LNT.S.YDF.SKND..F.S.GIPV.R...

```

■相同性検索データベースに 16S rRNA データを追加

DDBJ が WWW と E-mail で提供している相同性検索サービスのうち、FASTA, BLAST, SSEARCH の検索対象 データベースに、DDBJ 定期リリースから16S rRNA 配列データを取り出したデータセットを追加しました。詳細は下記の通りです。

- データベース名：16S rRNA (Prokaryotes)
- 作成方法：
 - DDBJ 定期リリースの BCT ティジョンのフラットファイルを対象に、以下の条件を満たす配列を抽出する。(300bp 以下は除外)
 - Feature の Key が "rRNA"
 - product 行に "16S", "RRS", "SSU" が含まれる
 - その後、配列の strand チェックを行い、アンチセンス配列と思われるものを除外した。
- Anonymous FTP サイト：
 - ファイル名：[16S.seq.gz](#) (FASTA 形式)
 - [FTP を利用したリリースデータ取得ページ](#) からご利用いただけます。

■BLAST, PSI-BLAST がバージョンアップ

DDBJ が WWW と E-mail サイトで提供している 相同性検索サービスのうち、BLAST (WWW と E-mail) と PSI-BLAST (WWW のみ) のバージョンアップ (2.2.14 --> 2.2.15) を行ないました。"J", "O" を含むアミノ酸の検索にも対応しました。

■大量データの公開

DDBJ が登録を受け付け、2006年12月から2007年1月にかけて DDBJ/EMBL/GenBank 国際塩基配列データベースから公開した大量データは以下の通りです。

▶メダカ GSS 430,750 エントリの新規公開 2006.12.27

国立遺伝学研究所から登録されたメダカ (*Oryzias latipes*, strain Hd-rR) の GSS データ 430,750 エントリが公開されました。アクセッション番号は以下の通りです。これらは12月22日, 23日, 26日分の新着データとして DDBJ より公開されており、anonymousFTP サイトから一括取得が可能です。

参考 URL：<http://dolphin.lab.nig.ac.jp>
アクセッション番号：DH000001-DH430750 (430,750 entries)
anonymousFTP 一括取得：ファイル名 [Oryzias_latipes_GSS_061222_1.seq.gz](#)

▶Xenopus laevis EST 129,076 エントリの新規公開 2006.12.6

基礎生物学研究所から登録されたアフリカツメガエル (*Xenopus laevis* ならびに *Xenopus tropicalis*) の EST 129,076 エントリが公開されました。アクセッション番号は以下の通りです。これらは12月5日分の新着データとして DDBJ より公開されており、anonymousFTP サイトから一括取得が可能です。

注) このデータは公開後の12月18日に訂正がありました。下記は正しい内容です。12月18日以前にデータを anonymousFTP 一括取得された方は、お手数ですが、データを再取得し、旧ファイルは破棄してください。

参考 URL：<http://www.shigen.nig.ac.jp/xenopus/top.jsp>
生物名ならびにアクセッション番号：

- Xenopus tropicalis DC135518-DC165593 (30076 entries)
- Xenopus tropicalis DC165594-DC195410 (29817 entries)
- Xenopus laevis DC066335-DC195410 (129,076 entries)より上記を除く

発生段階の表記修正

誤：stage 10.5 --->
正：mixture of equivalent stages 10.5-15 (early gastrula to mid neurula)
DC066335-DC099907 (33573 entries)
DC099908-DC135517 (35610 entries)

anonymousFTP 一括取得：ファイル名

- [Xenopus_tropicalis_EST_061215_1.seq.gz](#)
- [Xenopus_tropicalis_EST_061215_2.seq.gz](#)
- [Xenopus_laevis_EST_061215_1.seq.gz](#)
- [Xenopus_laevis_EST_061215_2.seq.gz](#)

▶Xenopus laevis EST 66,334 エントリの新規公開 2006.12.5

東京大学から登録されたアフリカツメガエル (*Xenopus laevis*) の EST データ 66,334 エントリを公開しました。アクセッション番号は以下の通りです。これらは12月2日分の新着データとして DDBJ より公開されており、anonymousFTP サイトから一括取得が可能です。

参考 URL：<http://www.shigen.nig.ac.jp/xenopus/top.jsp>
アクセッション番号：DC000001-DC066334 (66,334 entries)
anonymousFTP 一括取得：ファイル名 [Xenopus_laevis_EST_061202_1.seq.gz](#)

■DDBJ リリース 68 完成

DDBJ が管理・収集している塩基配列データベースは、リリースとして定期的に年4回オンライン上で公開しています。12月28日に DDBJ リリース68 を完成しました。リリース68 のエントリ数は 64,267,978、総塩基数は 68,259,314,742 です。通信環境の向上に合わせて、既に一部のファイルではファイルサイズが変更になっていますが、今回のリリースより全てのファイルサイズの最大が 1.5 GB に変更になりました。
[FTP による定期リリースおよび新着データのダウンロードサイト](#)

■統計 "Journal Ranking" について

DDBJ では、リリース65 より、統計データ「Journal ランキング上位100」を公開しています。この統計は、DDBJ リリースより Journal の情報を抜き出して集計したもので、同一雑誌を持つエントリを集計したものの [エントリ数] と、雑誌が同一で論文 (JOURNAL の値) が異なる件数 (ユニークな論文数) を集計したものの [論文数] とがあります。(Journal の表記方法には変更がありますが、リリースでは集計時の雑誌名を標準化する作業を行い、もの名のないようにしています。) これらの統計は、エントリ数や論文数から登録・投稿状況を見るところ、GenBankやEMBL にはない DDBJ 独自の統計データです。

- 論文数
http://www.ddbj.nig.ac.jp/breakdown_stats/journal-num.html
- エントリ数
http://www.ddbj.nig.ac.jp/breakdown_stats/journal-ent.html

学術雑誌に論文を投稿する際に国際塩基配列データベースのアクセッション番号の取得は、必要不可欠となりつつあります。論文を投稿する際の参考にするなど Journal Ranking の統計をどうぞご利用下さい。

■DDBJ スタッフロム19

「保存への努力」
小菅 武英 (DDBJアナター)

先日、近くの美術館で「よみがえる源氏物語絵巻展」が開催されていたので足を運んできた。国宝源氏物語絵巻といえば、12世紀ごろに作成されたといわれ、現存する絵巻物の中では最古のものだそうです。およそ900年前の絵巻物であるから、元の色彩を殆ど感じることができないほどに劣化している。これを、現在の非破壊検査技術を用いて、オランダの色彩、色使いなどを復元しようというのだ。つまり、蛍光X線分析装置を用いた絵巻物で使用されている金属元素を検出し、それを元で使用されている顔料、あるいは、染料を特定する。または、紫外線の絵巻物に照射することで、通常の見ええない部分を浮かび上げられるといった技法で、オリジナルの絵巻物を復元しようとする取り組みであった。復元された絵巻物の色彩は実に豊かであり、感慨深いものであった。

もちろん、科学的な検査で使用されている顔料や染料などが特定できたとしても、復元画家の力がなくてはそれを実際に復元するのはできないことである。もし、日本画の技法が900年の間に減っていたとしたら、科学的な分析のみで終わってしまい、価値のある復元を行うことは出来なかったであろう。

天災や戦火を経て、過去の資産を現在に伝えることができたのは、文化的な価値を認めた上で後世に伝えようとした人々の努力のたまものである。また、その技法を伝えることのできる日本画家を育て続けてきたことにも意味がある。

絵巻物の色彩のように、自然界の法則ではデータは劣化をしていくものである。DDBJでは、日々、大量の塩基配列データを蓄積し続けている。過去の塩基配列データを含め、すべては人間の努力から得られた資産とも考えられる。900年前は和紙に書いて保存していたものが現在では磁気ディスクに変わり、電子化によりデータ蓄積のスピードは常に増大し続けている。現在の科学が生んだ記憶メディアに保存された遺伝子データが、何世紀も単位で安定に維持できるとは言いわけではない。遺伝子データベースを、人類の永遠の資産として存続させ、誰でも利用可能な状態にしておくための努力を続けていかなければ、どこかの時代で消滅してしまうであろう。

現在の科学活動が生んだ遺伝子データベースが文化遺産のように千年単位で永続できるかどうかは分らないが、永久に保存する価値があるものとして、維持を続ける努力を惜しまないよう、教育活動の面でも援助が必要だとつくづく感じる。