

DDBJのサービス紹介

国立遺伝学研究所 大量遺伝情報研究室
中村保一

○|_| ゴメンナサイ

遺伝研スーパーコンピュータシステムの全面更新が 2012.3.1 (昨日！) に行われました。DDBJの運用系も含めて全面的にマシンが入れ替わりましたので、現在、いくつかのサービスが新システムへの移行途上のため休止中あるいは潜在的な不具合を抱えている（かもしれない）状況です。

ですので...

DDBJ等のサービス紹介

国立遺伝学研究所 大量遺伝情報研究室
中村保一

自己紹介

Google 検索 約 52,900 件 (0.18 秒)

すべて 他のキーワード: 中村保一twitter

画像 中村保一 | ynlab@nig
charles.genes.nig.ac.jp/members/yn/ - キャッシュ

地図 誰. なかむらやすかず 博士 (理学) 国立遺伝学研究所 大量遺伝情報研究室 教授.

動画 Curriculum Vitae. モットー：費用対効果の追求問い合わせ：あなたの真摯な努力はみんなの迷惑、かもよ行動規範：誠意をもってテキトーに、研究／業務。研究領域：ゲノム生物学 ...

ニュース

ショッピング

もっと見る 大量遺伝情報研究室・中村研究室 - 国立遺伝学研究所<研究・組織>
www.nig.ac.jp/section/yn/yn-j.html - キャッシュ

静岡県三島市 場所を変更 生物研究の基盤データベースとしてのDDBJ事業の推進. 教授 中村 保一 yanakamu 助教 神沼 英里 ekaminum. 研究室URL : http://charles.genes.nig.ac.jp/. 高速かつ大量に決定される塩基配列情報は、生物学のあらゆる分野で活用される情報基盤です。

ウェブ全体から検索 ynlab@nig | Genome Informatics laboratory, CIB-DDBJ, National ...
charles.genes.nig.ac.jp/ - キャッシュ

日本語のページを検索 翻訳して検索 もっとツールを見る 5日前 - 応募者は略歴(学歴・職歴・資格)をEメールにて中村保一（教授）へご送付下さい。書類選考の上、面接をして採用を決めます。適任者が決まり次第、募集を締め切ります。 国立遺伝学研究所 大量遺伝情報研究室静岡県三島市谷田1111 tel: ...

中村保一 博士 (猫) (@yaskaz) on Twitter
twitter.com/#/yaskaz - キャッシュ

Sign up for Twitter to follow 中村保一 博士 (猫) (@yaskaz). 猫教授。生物学者、UNIX使いのMac愛好家、二児の父で無類の猫好き。

⌚ 超略歴

-2008 かずきDNA研 室長／東大新領域 情報生命 客員准教授

2009- 遺伝研 大量遺伝情報研究室 教授

45歳、猫好き

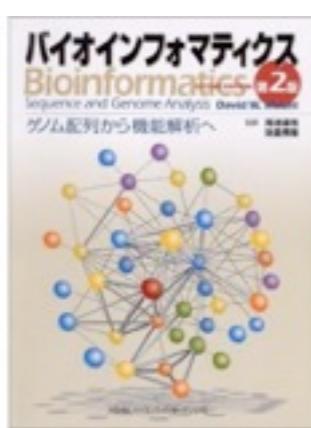
主要な著作・翻訳



<http://www.amazon.co.jp/dp/4758108110>
編集代表



<http://www.amazon.co.jp/dp/4897068746>
企画・編集



<http://www.amazon.co.jp/dp/4895924262>
翻訳分担（第9章・遺伝子予測と遺伝子調節）

【Trad】

DDBJ

DNA Data Bank of Japan

DDBJ (<http://www.ddbj.nig.ac.jp/>)

DDBJ
DNA Data Bank of Japan

ENGLISH  

HOME 塩基配列の登録 利用の手引き 検索・解析 FTP・WebAPI レポート・統計 お問い合わせ サイト内検索

▶ DDBJの紹介
▶ Q&A集
▶ 塩基配列の登録
▶ SAKURA
▶ 大量登録システム(MSS)
▶ データの修正・更新
▶ DDBJ Sequence Read Archive
▶ DDBJ Trace Archive
▶ プロジェクトの登録
▶ DDBJ BioProject Database
▶ 検索
▶ getentry
▶ ARSA
▶ TXSearch
▶ BLAST
▶ 系統解析
▶ ClustalW
▶ NGSデータ解析

DDBJ : DNA Data Bank of Japan

DDBJ（日本DNAデータバンク）は欧州と米国の対応機関（EBIおよびNCBI）と密接に協力しながら DDBJ/EMBL/GenBank 国際塩基配列データベースを構築している三大国際DNAデータバンクのひとつです


Photo by Hideki Nagasaki

▶ 一覧へ

Hot Topics

- 2012.02.13 [DDBJ エントリへのリンク設定方法の変更](#)
- 2012.02.13 [DDBJ の新しい キーワード・エントリ検索システムについて](#)
- 2012.02.02 [ユーカリ \(*Eucalyptus camaldulensis*\) EST データの公開](#)

▶ 一覧へ

Maintenance

- 2012.02.20 [DDBJ サービスの中止・変更について \(2/15 現在の状況\)](#)
- 2012.02.14 (2/25) [国立遺伝学研究所ならびに DDBJ ネットワークの中止](#)
- 2012.02.13 (再開) (2012/2/23-27) [SAKURA によるデータ受付の一時休止](#)

▶ 一覧へ

Information

- [2012/3実施のコンピュータシステム移行に伴うお知らせ \(一覧\)](#)
- [DDBJ メールマガジン No.69 配信](#)

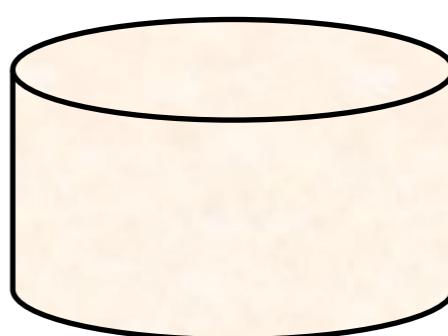
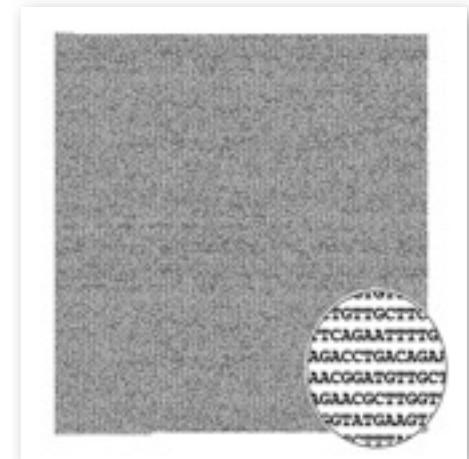
塩基配列の登録・更新  FTP・Web API 

遺伝研／DDBJは静岡県三島市にあります



塩基配列データバンク

- 全世界で解読された塩基配列情報を
査定して受入れ
- データベースに蓄積し
- 公開して共有する



国際塩基配列データベースの一員

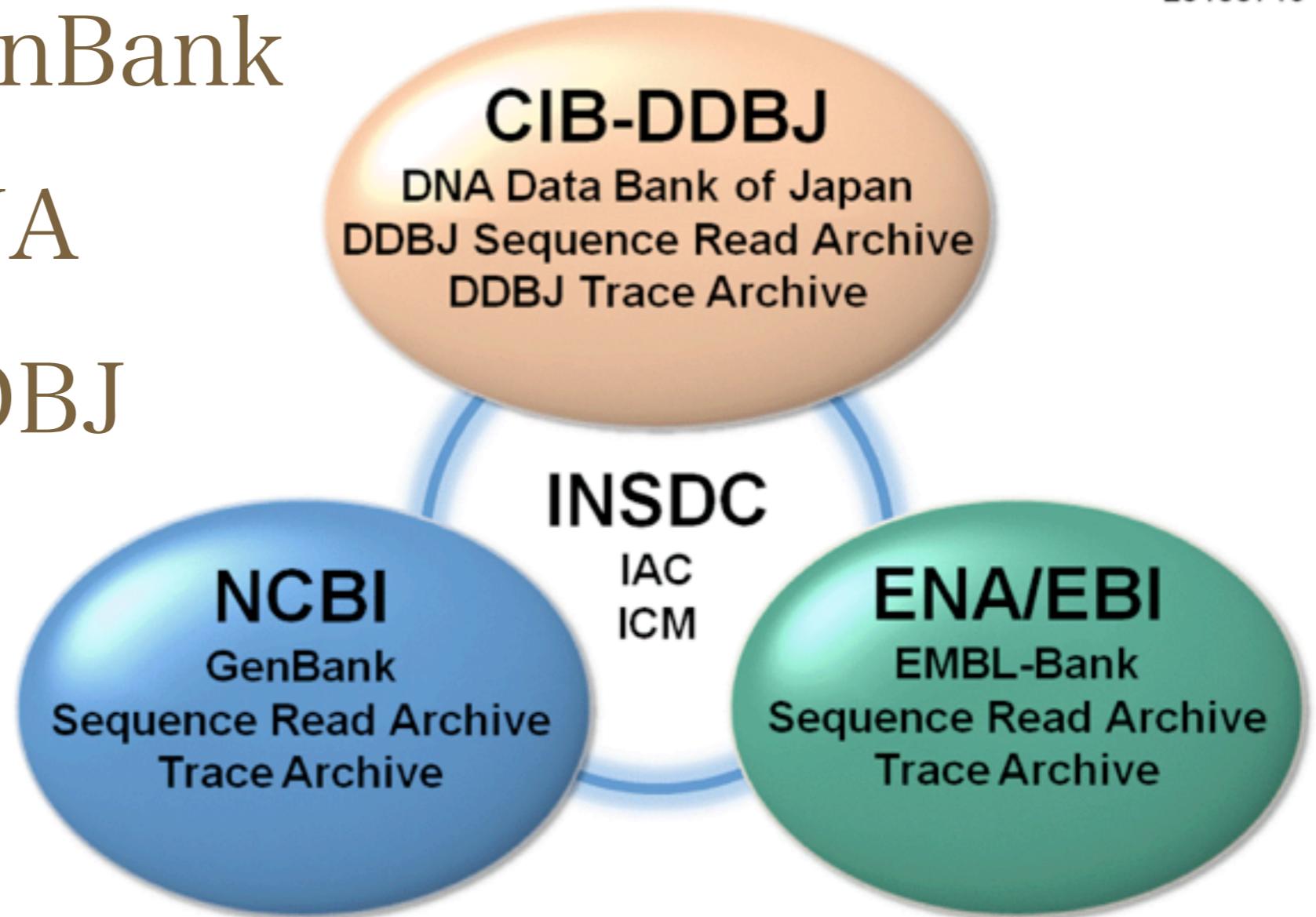
International Nucleotide Sequence Databank Collaboration

20100715

米国: GenBank

欧洲: ENA

日本: DDBJ





articles

Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*

The *Arabidopsis* Genome Initiative*

* Authorship of this paper should be cited as 'The *Arabidopsis* Genome Initiative'. A full list of contributors appears at the end of this paper

The flowering plant *Arabidopsis thaliana* is an important model system for identifying genes and determining their functions. Here we report the analysis of the genomic sequence of *Arabidopsis*. The sequenced regions cover 115.4 megabases of the 125-megabase genome and extend into centromeric regions. The evolution of *Arabidopsis* involved a whole-genome duplication, followed by subsequent gene loss and extensive local gene duplications, giving rise to a dynamic genome enriched by lateral gene transfer from a cyanobacterial-like ancestor of the plastid. The genome contains 25,498 genes encoding proteins from 11,000 families, similar to the functional diversity of *Drosophila* and *Caenorhabditis elegans*—the other sequenced multicellular eukaryotes. *Arabidopsis* has many families of new proteins but also lacks several common protein families, indicating that the sets of common proteins have undergone differential expansion and contraction in the three multicellular eukaryotes. This is the first complete genome sequence of a plant and provides the foundations for more comprehensive comparison of conserved processes in all eukaryotes, identifying a wide range of plant-specific gene functions and establishing rapid systematic ways to identify genes for crop improvement.

The plant and animal kingdoms evolved independently from unicellular eukaryotes and represent highly contrasting life forms. The genome sequences of *C. elegans*¹ and *Drosophila*² reveal that metazoans share a great deal of genetic information required for developmental and physiological processes, but these genome sequences represent a limited survey of multicellular organisms. Flowering plants have unique organizational and physiological properties in addition to ancestral features conserved between plants and animals. The genome sequence of a plant provides a means for understanding the genetic basis of differences between plants and other eukaryotes, and provides the foundation for detailed functional characterization of plant genes.

Arabidopsis thaliana has many advantages for genome analysis, including a short generation time, small size, large number of offspring, and a relatively small nuclear genome. These advantages promoted the growth of a scientific community that has investigated the biological processes of *Arabidopsis* and has characterized many genes³. To support these activities, an international collaboration (the *Arabidopsis* Genome Initiative, AGI) began sequencing the genome in 1996. The sequence of chromosomes 2 and 4 have been reported^{4,5}, and the accompanying Letters describe the sequences of chromosomes 1 (ref. 6), 3 (ref. 7) and 5 (ref. 8).

Here we report analysis of the completed *Arabidopsis* genome sequence, including annotation of predicted genes and assignment of functional categories. We also describe chromosome dynamics and architecture, the distribution of transposable elements and other repeats, the extent of lateral gene transfer from organelles, and the comparison of the genome sequence and structure to that of other *Arabidopsis* accessions (distinctive lines maintained by single-seed descent) and plant species. This report is the summation of work by experts interested in many biological processes selected to illuminate plant-specific functions including defence, photomorphogenesis, gene regulation, development, metabolism, transport and DNA repair.

The identification of many new members of receptor families, cellular components for plant-specific functions, genes of bacterial origin whose functions are now integrated with typical eukaryotic components, independent evolution of several families of transcription factors, and suggestions of as yet uncharacterized metabolic pathways are a few more highlights of this work. The implications of these discoveries are not only relevant for plant

biologists, but will also affect medical science, evolutionary biology, bioinformatics, environmental biology, functional and comparative genomics, and molecular medicine.

Overview of sequencing strategy

We used large-insert clones of centromeric chromosomal DNA (BAC), phage (P1) and yeast artificial chromosome (YAC) libraries^{9–11} as the primary strategy for genome assembly. Early stages of genome sequencing were guided by physical maps of the genome obtained by linkage analysis¹² or by restriction fragment length polymorphism (RFLP) analysis¹³ or hybridization¹⁴ or polymerase chain reaction (PCR) of cloned sites and by hybridization of genomic DNA with the genetic map and provided a foundation for assembling sets of contigs into sequence-ready tiling paths. End sequence (http://www.tigr.org/tdb/at/abe/bac_end_search.html) of 47,788 BAC clones was used to extend contigs from BACs anchored by marker content and to integrate contigs.

Ten contigs representing the chromosome arms and centromeric heterochromatin were assembled from 1,569 BAC, TAC, cosmid and P1 clones (average insert size 100 kilobases (kb)). Twenty-two PCR products were amplified directly from genomic DNA and sequenced to link regions not covered by cloned DNA or to optimize the minimal tiling path. Telomere sequence was obtained from specific yeast artificial chromosome (YAC) and phage clones, and from inverse polymerase chain reaction (IPCR) products derived from genomic DNA. Clone fingerprints, together with BAC end sequences, were generally adequate for selection of clones for sequencing over most of the genome. In the centromeric regions, these physical mapping methods were supplemented with genetic mapping to identify contig positions and orientation¹⁵.

Selected clones were sequenced on both strands and assembled using standard techniques. Comparison of independently derived sequence of overlapping regions and independent assembly of sequenced clones revealed accuracy rates between 99.99 and 99.999%. Over half of the sequence differences were between genomic and BAC clone sequence. All available sequenced genetic markers were integrated into sequence assemblies to verify sequence contigs^{16,17}. The total length of sequenced regions, which extend from either the telomeres or ribosomal DNA repeats to the 180-base-pair

学術誌

前編

遺伝子に関する論文には「登録」が不可欠

論文投稿時の注意: 論文の著者は、論文で言及した塩基配列については、原稿の末尾にデータ公開欄を設け、インターネットで参照が可能な公共データベースの登録番号を掲載しなければならない

The screenshot shows the homepage of the journal 'nature'. At the top, there's a banner with the text 'From across nature.com' and 'access the site now!'. Below the banner, the 'nature' logo is prominently displayed, followed by the subtitle 'International weekly journal of science'. A search bar at the top right includes 'Search this journal' and 'go Advanced search' buttons. On the left side, there's a sidebar with 'Journal content' links such as 'Journal home', 'Advance online publication', 'Current issue', 'Nature News', 'Archive', 'Supplements', 'Web focuses', 'Podcasts', 'Videos', and 'News Specials'. Another section titled 'Journal information' includes links for 'About the journal' and 'For authors'. The main content area features a large heading 'Formatting guide: manuscript preparation and submission'. Below this, there's a section titled 'Information Sheets for Downloading' with a list of downloadable files, including 'Manuscript preparation and submission (doc 103KB)', 'Section summaries (doc 40KB)', 'Communications Arising (doc 60.5KB)', 'Annotated example: summary paragraph for Letters (doc 40KB)', 'Annotated example: end notes (doc 225KB)', 'Statistical checklist (doc 44KB)', 'Characterization of chemical materials (doc 43KB)', and 'Nature and Nature Chemical Biology style for chemical structures (PDF 243KB)'. There's also a link to 'See the full list of information sheets'. Further down, there's a note about the guide describing how to prepare contributions for submission, mentioning a short version available above and a full version below for new contributors. A warning states that failure to adhere to these guidelines can seriously delay the handling of your contribution. To the right, there's a sidebar with a 'nature' logo and a '30% Spring discount to Nature!' offer, along with links for 'Sign up for e-alerts', 'Recommend to your library', and 'Web feed'. Another section titled 'open innovation challenges' lists 'Mechanisms of Action of Preservative Boosters in Bacteria' with a deadline of Jun 14 2010 and a reward of \$20,000 USD, and 'Long-term Preservation of Multicellular Organisms in a...'.

DDBJ登録ファイルの例

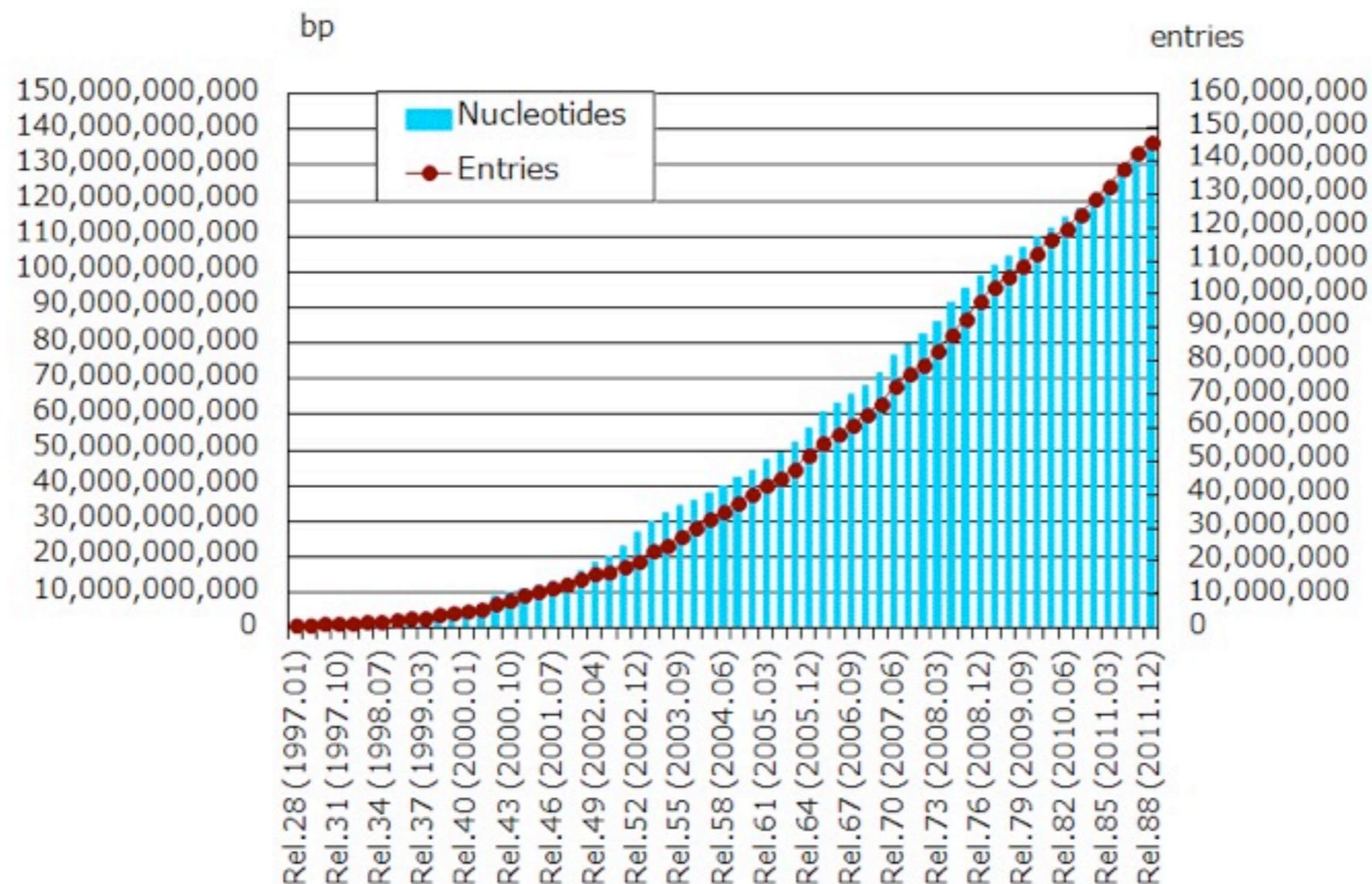
LOCUS HUMIL2HOM 397 bp DNA linear HUM 27-APR-1993
DEFINITION Human interleukin 2 (IL-2)-like DNA.
ACCESSION M13784
VERSION M13784.1
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 397)
AUTHORS Mita,S., Maeda,S. and Shimada,K.
TITLE Characterization of human genomic DNA sequences homologous to the
interleukin 2 cDNA
JOURNAL Biochem. Biophys. Res. Commun. 138 (2), 966-973 (1986)
PUBMED 3017347
COMMENT Original source text: Human placenta DNA, clone Lm HoIL2-3.
Numerous stop codons are found in the interleukin 2-like IIa DNA.
FEATURES Location/Qualifiers
source 1..397
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 117 a 84 c 48 g 148 t
ORIGIN RsaI site.
1 actgatttat tttaataaaa attacaagag atttaattt taaacccaaa agttctttta
61 ttgcatctca ctgtgttag ctttgttac ctttgagaa ggcctgagat aataacttgc
121 ttcttcact ctttcatcag ctcctgtaac ctttttcct tagttctta actgatgttg
181 tggcctgctg ctaaaaacgc tttatctaa agttctaaaa ggaaatgttt tcattctaaca
241 taacattctg ggctcttgac tttatgaaat caaaaacttt cacttatgac caggatacac
301 tcttcctctg tctaactaat tcaagcacta tcttcattca ttttgacttg cagattatcc
361 aaacagactc cccataatga aaagcaatca cactgca
//

現在の塩基配列データの量

塩基数: 1,400億

登録数: 1.5億

DDBJ/EMBL/GenBank database growth



Note: CON division is not counted in statistics of DDBJ periodical

DDBJに多くの配列が登録された生物種

Images created by the Wordle.net web application are licensed under a Creative Commons Attribution 3.0 United States License.



DDBJに登録されている生物種 Top 100の
ワードクラウド

キーワード検索あれこれ

- ⌚ ARSA [DDBJ]: <http://arsa.ddbj.nig.ac.jp>
- ⌚ Entrez [汎用]: <http://www.ncbi.nlm.nih.gov> から
- ⌚ Google Scholar [学術資料一般]: <http://scholar.google.com>
- ⌚ GOLD [ゲノムプロジェクト]: <http://www.genomesonline.org>

いきなり遺伝子のIDや一般名称でgoogle検索する、という手もあります

DDBJ (<http://www.ddbj.nig.ac.jp/>)

DDBJ
DNA Data Bank of Japan

ENGLISH  

HOME 塩基配列の登録 利用の手引き 検索・解析 FTP・WebAPI レポート・統計 お問い合わせ サイト内検索

▶ DDBJの紹介
▶ Q&A集
▶ 塩基配列の登録
SAKURA
大量登録システム(MSS)
データの修正・更新
DDBJ Sequence Read Archive
DDBJ Trace Archive
▶ プロジェクトの登録
DDBJ BioProject Database
▶ 検索
entry
ARSA (Red circle)
TXSearch
BLAST
▶ 系統解析
ClustalW
▶ NGSデータ解析

DDBJ : DNA Data Bank of Japan

DDBJ（日本DNAデータバンク）は欧州と米国の対応機関（EBIおよびNCBI）と密接に協力しながら DDBJ/EMBL/GenBank 国際塩基配列データベースを構築している三大国際DNAデータバンクのひとつです


Photo by Hideki Nagasaki

Hot Topics ▶ 一覧へ

- 2012.02.13 [DDBJ エントリへのリンク設定方法の変更](#)
- 2012.02.13 [DDBJ の新しい キーワード・エントリ検索システムについて](#)
- 2012.02.02 [ユーカリ \(*Eucalyptus camaldulensis*\) EST データの公開](#)

Maintenance ▶ 一覧へ

- 2012.02.20 [DDBJ サービスの中止・変更について \(2/15 現在の状況\)](#)
- 2012.02.14 [\(2/25\) 国立遺伝学研究所ならびに DDBJ ネットワークの中止](#)
- 2012.02.13 [\(再開\) \(2012/2/23-27\) SAKURA によるデータ受付の一時休止](#)

Information

- [2012/3実施のコンピュータシステム移行に伴うお知らせ \(一覧\)](#)
- [DDBJ メールマガジン No.69 配信](#)

塩基配列の登録・更新  FTP・Web API 

ARSA

<http://arsa.ddbj.nig.ac.jp/>

The screenshot shows a web browser window for the ARSA (All-round Retrieval of Sequence and Annotation) search engine. The URL in the address bar is <http://arsa.ddbj.nig.ac.jp/html/>. The page title is "ARSA All-round Retrieval of Sequence and Annotation". Below the title, there are two search buttons: "DDBJQuick Search" and "DDBJAdvanced Search". The "DDBJQuick Search" button is active. In the search input field, the text "E. coli O157" is entered. A red oval highlights this input field. To the right of the input field is a "Search" button. Below the search form, a message in Japanese states: "検索条件を複数入力する場合は、&(AND条件)、|(OR条件)、!(AND NOT条件)を指定することができます。" (When entering multiple search conditions, you can specify &(AND condition), |(OR condition), !(AND NOT condition)). At the bottom of the search form, it says "Copyright© DDBJ All rights reserved." A large text overlay "E. coli O157 を検索" is displayed prominently in the center of the page.

検索結果: 絞り込みが必要だな...

The screenshot shows a web browser window with the following details:

- Address Bar:** QuickSearch, arsa.ddbj.nig.ac.jp/html/
- Toolbar:** Back, Forward, Stop, Refresh, Home, Bookmarks, History, etc.
- Menu Bar:** W, DBCLS, E, P, R, S, Papers, Tech, travel, Dilgolet, その他ブックマーク
- Title Bar:** ARSA All-round Retrieval of Sequence and Annotation
- Search Bar:** QuickSearch (selected), DDBJAdvanced Search, Search button (with input field containing "E. coli O157")
- Help Text:** 検索条件を複数入力する場合は、&(AND条件)、| (OR条件)、! (AND NOT条件)を指定することができます。
- Download Options:** FlatFile (selected), XML, fasta, View, Full Download, All Select
- Table Headers:** PrimaryAccessionNumber, Definition, moltype, Organism, Length
- Table Data:** Three rows of results for Escherichia coli O157 strains.

PrimaryAccessionNumber	Definition	moltype	Organism	Length
<input type="checkbox"/> AB011548	Escherichia coli O157:H7 str. Sakai plasmid pOSAK1 DNA, complete sequence.	DNA	Escherichia coli O157:H7 str. Sakai	3306
<input type="checkbox"/> AB011549	Escherichia coli O157:H7 str. Sakai plasmid pO157 DNA, complete sequence.	DNA	Escherichia coli O157:H7 str. Sakai	92721
	Escherichia coli O157:H7 hemG, rrsA, ileT, alaT, rrlA, rrfA, mobB, mobA genes for protoporphyrin oxidase			

Left Sidebar: A sidebar lists various search terms and their counts: DDBJ(2136), Patent_AA(3), Escherichia coli(680), Unknown_(561), Escherichia coli_O157:H7(241), unidentified(167), synthetic(160), construct(160), Escherichia coli_O157:H7 str.(130), EDL933(130), Pasteuria penetrans(15), Gallus gallus(10), Hydractinia echinata(8).

Advanced Searchへ

The screenshot shows the ARSA (All-round Retrieval of Sequence and Annotation) search interface. At the top, there is a navigation bar with links to various sections like W, DBCLS, E, P, R, S, Papers, Tech, travel, and Dilgolet. Below the navigation bar is a blue header with the ARSA logo and the text "All-round Retrieval of Sequence and Annotation". The main content area has two tabs: "DDBJQuick Search" and "DDBJAdvanced Search", with "DDBJAdvanced Search" circled in red. A search bar below the tabs contains the query "E. coli O157". A message below the search bar states: "検索条件を複数入力する場合は、&(AND条件)、| (OR条件)、!(AND NOT条件)を指定することができます。" To the left of the search results, there is a sidebar with a list of search terms: DDBJ(2136), Patent_AA(3), Escherichia coli(680), Unknown_(561), Escherichia coli_O157:H7(241), unidentified(167), synthetic construct(160), Escherichia coli_O157:H7 str.(130), EDL933(130), Pasteuria penetrans(15), Gallus gallus(10), Hydractinia echinata(8). The main results table has columns: PrimaryAccessionNumber, Definition, moltype, Organism, and Length. The first two rows show entries for Escherichia coli O157 strains, while the third row shows a partial entry for Hydractinia echinata.

PrimaryAccessionNumber	Definition	moltype	Organism	Length
<input type="checkbox"/> AB011548	Escherichia coli O157:H7 str. Sakai plasmid pOSAK1 DNA, complete sequence.	DNA	Escherichia coli O157:H7 str. Sakai	3306
<input type="checkbox"/> AB011549	Escherichia coli O157:H7 str. Sakai plasmid pO157 DNA, complete sequence.	DNA	Escherichia coli O157:H7 str. Sakai	92721
	Escherichia coli O157:H7 hemG, rrsA, ileT, alaT, rrlA, rrfA, mobB, mobA genes for protoporphyrin oxidase			

いくつかの特徴で絞り込み

AdvancedSearch arsa.ddbj.nig.ac.jp/html/AdvancedSearchMenu

W DBCLS E P R S Papers Tech travel Dilgolet その他のブックマーク

Combine Searches with **&(AND)**

All Text =
Accession Number =
Primary Accession Number =
Division = BCT CON ENV HTC HTG HUM INV
MAM PAT PHG PLN PRI ROD STS
SYN TSA UNA VRL VRT

Sequence Length = -

Molecular

Type DNA RNA cRNA mRNA rRNA tRNA
Form circular linear

Date = -
Definition =
Comment =
Keyword =
Organism = E. coli
Taxon =

おやおや

AdvancedSearch × arsa.ddbj.nig.ac.jp/html/AdvancedSearch ☆ 🔍

W DBCLS E P R S Papers Tech travel Dlgolet その他のブックマーク

ARS A All-round Retrieval of Sequence and Annotation

DDBJQuick Search DDBJAdvanced Search Refine Search

DDBJ(3)

FlatFile XML fasta View Full Download

All Select

PrimaryAccessionNumber	Definition	moltype	Organism	Length
<input type="checkbox"/> AB602479	C. glutamicum-E. coli shuttle vector pCRB12 DNA, complete sequence.	DNA	C. glutamicum-E. coli shuttle vector pCRB12	4569
<input type="checkbox"/> HM126493	C. glutamicum-E. coli shuttle vector pCRB62, complete sequence.	DNA	C. glutamicum-E. coli shuttle vector pCRB62	5914
<input type="checkbox"/> HM126494	C. glutamicum-E. coli shuttle vector pCRB12, complete sequence.	DNA	C. glutamicum-E. coli shuttle vector pCRB12	4569

[1]

Copyright© DDBJ All rights reserved.

Escherichia coli O157 sakai でどや

AdvancedSearch arsa.ddbj.nig.ac.jp/html/AdvancedSearchMenu

W DBCLS E P R S Papers Tech travel Digolet その他のブックマーク

All Text Accession Number Primary Accession Number
Division BCT CON ENV HTC HTG HUM INV
 MAM PAT PHG PLN PRI ROD STS
 SYN TSA UNA VRL VRT

Sequence Length = =

Molecular

Type DNA RNA cRNA mRNA rRNA tRNA
Form circular linear

Date = Definition
Comment
Keyword
Organism Escherichia coli O157 sakai
Taxon

Reference

Escherichia coli O157 Sakai 株ゲノム

The screenshot shows a web browser window for the ARSA (All-round Retrieval of Sequence and Annotation) search interface. The URL is arsa.ddbj.nig.ac.jp/html/AdvancedSearch. The search results are displayed for *Escherichia coli* O157 str. Sakai.

Search parameters:

- DDBJ(3)
- Format: FlatFile (selected)
- View
- Full Download
- All Select
- Refine Search

Table of search results:

PrimaryAccessionNumber	Definition	moltype	Organism	Length
AB011548	Escherichia coli O157:H7 str. Sakai plasmid pOSAK1 DNA, complete sequence.	DNA	Escherichia coli O157:H7 str. Sakai	3306
AB011549	Escherichia coli O157:H7 str. Sakai plasmid pO157 DNA, complete sequence.	DNA	Escherichia coli O157:H7 str. Sakai	92721
BA000007	Escherichia coli O157:H7 str. Sakai DNA, complete genome.	DNA	Escherichia coli O157:H7 str. Sakai	5498450

BA00007を開く ⇒ フラットファイル閲覧

The screenshot shows a web browser window with a green header bar. The title bar says "AdvancedSearch" and "flatfile". The address bar shows the URL "getentry.ddbj.nig.ac.jp/getentry/ddbj/BA000007?filetype=html". The toolbar below the address bar includes icons for W, DBCLS, E, P, R, S, Papers, Tech, travel, and Dilgolet, along with a "その他ブックマーク" (Other Bookmarks) icon.

The main content area displays the following flatfile data for DDBJ entry BA000007:

LOCUS BA000007 5498450 bp DNA circular BCT 18-JAN-2008
DEFINITION Escherichia coli O157:H7 str. Sakai DNA, complete genome.
ACCESSION [BA000007](#) AP002550-AP002569
VERSION BA000007.2
DBLINK BioProject:PRJNA226
KEYWORDS .
SOURCE Escherichia coli O157:H7 str. Sakai
ORGANISM [Escherichia coli O157:H7 str. Sakai](#)
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 5498450)
AUTHORS Hattori,M., Ishii,K. and Shiba,T.
TITLE Direct Submission
JOURNAL Submitted (26-JUN-2000) to the DDBJ/EMBL/GenBank databases.
Contact:Masahira Hattori
Graduate School of Frontier Sciences, University of Tokyo; 5-1-5
Kashiwanoha, Kashiwa, Chiba 277-8561, Japan
REFERENCE 2
AUTHORS Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
Yokoyama,K., Han,C., Ohtsubo,E., Nakayama,K., Murata,T.,
Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
Shinagawa,H.
TITLE Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12.
JOURNAL DNA Res. 8, 11-22 (2001)
REFERENCE 3
AUTHORS Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
Sasakawa,C. and Shinagawa,H.
TITLE Complete nucleotide sequence of the prophage VT2-Sakai carrying
the verotoxin 2 genes of the enterohemorrhagic Escherichia coli

Entrez <http://www.ncbi.nlm.nih.gov/Entrez>

↓ 【コツ】コレを打ち込むより、googleで「Entrez」検索した結果から開くほうが速いです

PubMedも包含する、NCBIの生物系検索決定版

The screenshot shows the NCBI Entrez homepage. At the top, there's a green banner with the NCBI logo and the text "Entrez cross-database search". Below it is a navigation bar with links for W, DBCLS, E, P, R, S, Papers, Tech, travel, and Dligolet, along with a "その他" (Other) link. The main content area features the NCBI logo and the text "Entrez, The Life Sciences Search Engine". A search bar at the top says "Search across databases" with "GO" and "Clear" buttons. Below this, a section titled "Welcome to the Entrez cross-database search page" lists various databases: PubMed, PubMed Central, Site Search, Nucleotide, EST, GSS, Books, OMIM, dbGaP, UniGene, CDD, and Protein. Each database entry includes a small icon and a brief description.

Entrez cross-database search

www.ncbi.nlm.nih.gov/sites/gquery

W DBCLS E P R S Papers Tech travel Dligolet その他

NCBI

Entrez, The Life Sciences Search Engine.

HOME SEARCH SITE MAP PubMed All Databases Human Genome GenBank Map Viewer BLA

Search across databases GO Clear Help

Welcome to the Entrez cross-database search page

PubMed: biomedical literature citations and abstracts	Books: online books
PubMed Central: free, full text journal articles	OMIM: online Mendelian Inheritance in Man
Site Search: NCBI web and FTP sites	

Nucleotide: Core subset of nucleotide sequence records	dbGaP: genotype and phenotype
EST: Expressed Sequence Tag records	UniGene: gene-oriented clusters of transcript sequences
GSS: Genome Survey Sequence records	CDD: conserved protein domain database

(cc) BY © 2012 中村保一 (国立遺伝学研究所) licensed under CC表示2.1日本

Entrezで同じ文字列を検索

The screenshot shows the NCBI Entrez search interface. A red circle highlights the search term "Escherichia coli O157 sakai" in the search bar. Another red circle highlights the "Genome" database entry in the results table.

Search results for "Escherichia coli O157 sakai":

Database	Count	Description
PubMed	101	biomedical literature citations and abstracts
PubMed Central	223	free, full text journal articles
Site Search	4	NCBI web and FTP sites
Nucleotide	17606	Core subset of nucleotide sequence records
EST	4	Expressed Sequence Tag records
GSS	none	Genome Survey Sequence records
Protein	31026	sequence database
Genome	1	whole genome sequences
Structure	2	three-dimensional macromolecular structures
Taxonomy	none	organisms in GenBank
dbGaP	2	genotype and phenotype
UniGene	1	gene-oriented clusters of transcript sequences
CDD	2	conserved protein domain database
Clone	none	integrated data for clone resources
UniSTS	none	markers and mapping data
PopSet	3591	population study data sets
GEO Profiles	none	expression and molecular abundance profiles
SEO DataSets	none	experimental sets of GEO data

Entrez: 大腸菌ゲノムページ

Escherichia coli (ID 167) - www.ncbi.nlm.nih.gov/genome/?term=Escherichia%20coli%20O157%20sakai

W DBCLS E P R S Papers Tech travel Dilgolet その他のブックマーク

NCBI Resources How To My NCBI Sign In

Genome Genome Escherichia coli O157 sakai Search Save search Limits Advanced Help

Display Settings: Overview Send to:

Escherichia coli
A well-studied enteric bacterium

Lineage: Bacteria[3002]; Proteobacteria[1251]; Gammaproteobacteria[564]; Enterobacteriales[147]; Enterobacteriaceae[147]; Escherichia[16]; Escherichia coli[2]

Escherichia coli. This organism was named for its discoverer, Theodore Escherich, and is one of the premier model organisms used in the study of bacterial genetics, physiology, and biochemistry. This enteric organism is typically present in the lower intestine of humans, where it is the dominant facultative anaerobe present, but it is [More...](#)

Organism Overview See also: [Genome list](#) [Plasmid list](#)

Sub-species tree

Genome Projects

Highest level of Assembly	RefSeq	Primary	All
Chromosomes	31	56	56
Scaffolds or contigs	60	267	267
SRA or Traces	-	67	67
No data	-	653	653
Total	91	1043	1043

Tools
BLAST Genome

Publications
DNA sequence analysis of the composite plasmid pTC conferring virulence and antimicrobial resistance for porcine enterotoxigenic *Escherichia coli*. [Int J Med Microbiol 2012]
Genome sequences and phylogenetic analysis of K88- and F18-positive porcine enterotoxigenic *Escherichia coli*. [J Bacteriol 2012]
Sequence of pR3521, an IncB plasmid from *Escherichia coli* encoding ACC-4, SCO-1, and TEM-1 beta-lactamases. [Antimicrob Agents Chemother 2011] See more....

Related information
BioProject
Gene
Protein Clusters
Components
Protein

(cc) BY © 2012 中村保一 (国立遺伝学研究所) licensed under CC表示2.1日本

Entrez: INSDC のリンクを開くと

Escherichia coli (ID 167) - Go Escherichia coli O157:H7 str. Sakai

www.ncbi.nlm.nih.gov/genome/167?project_id=57781

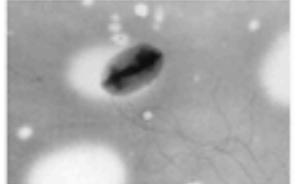
W DBCLS E P R S Papers Tech travel Dlgolet その他のブックマーク

NCBI Resources How To My NCBI Sign In

Genome Genome Search Limits Advanced Help

Display Settings: Overview Send to:

Return to Escherichia coli

 **Escherichia coli O157:H7 str. Sakai**
Enterohemorrhagic Escherichia coli

Lineage: Bacteria[3002]; Proteobacteria[1251]; Gammaproteobacteria[564]; Enterobacteriales[147]; Enterobacteriaceae[147]; Escherichia[16]; Escherichia coli[2]; Escherichia coli O157:H7[0]; Escherichia coli O157:H7 str. Sakai[0]

Escherichia coli O157:H7. This strain is associated with Hamburger disease, which is caused by the contamination of meat products by enterohemorrhagic *E. coli* (EHEC). The identifier O157:H7 refers to the serotype of EHEC, and reflects the specific antigenic markers found on the surface of the cell. EHEC attaches and effaces to cells [More...](#)

Genome details for Escherichia coli O157:H7 str. Sakai

See [Genome list](#)

Feature counts are from RefSeq where it is available

Escherichia coli O157:H7 str. Sakai See Protein Details

GIRC

Enterohemorrhagic Escherichia coli

Type	Name	RefSeq	INSDC	Size (Mb)	GC%	protein	rrna	trna	other rna	gene
Chr	-	NC_002695.1	BA000007.2	9.5	50.5	5,230	22	105	14	5,372
Plsm	pO157	NC_002128.1	AB011549.2	0.09	47.6	85	-	-	-	85
Plsm	pOSAK1	NC_002127.1	AB011548.2	0	43.4	3	-	-	-	3

Related Information

BioProject

Gene

Protein Clusters

Entrez のフラットファイル表示

The screenshot shows a web browser displaying the NCBI Nucleotide database. The URL in the address bar is www.ncbi.nlm.nih.gov/nuccore/47118301. The page title is "Escherichia coli O157:H7 str. Sakai DNA, complete genome". The search bar at the top indicates the query is for "Nucleotide". On the right side, there is a "Customize view" panel with options for "Basic Features" (radio button selected for "Gene, RNA, and CDS features only") and "Display options" (checkboxes for "Show sequence" and "Show reverse complement"). Below the main content area, there are sections for "Analyze this sequence" (Run BLAST, Pick Primers, Highlight Sequence Features) and "LinkOut to external resources" (Institute for Transcriptional Informatics, REBASE enzyme XfaMrrP). The main content area displays the genomic record for BA000007.2, including fields like LOCUS, DEFINITION, ACCESSION, VERSION, DBLINK, KEYWORDS, SOURCE, ORGANISM, and REFERENCE.

Entrez のフラットファイル表示

Escherichia coli (ID 167) - G Escherichia coli O157:H7 str. Sakai DNA, complete genome

www.ncbi.nlm.nih.gov/nuccore/47118301

W DBCLS E P R S Papers Tech travel Dilgolet その他のブックマーク

NCBI Resources How To My NCBI Sign In

Nucleotide Nucleotide Search Limits Advanced Help

Display Settings: GenBank Send: Change region shown

i Sequence not displayed. Use 'Customize View' section for control.

Escherichia coli O157:H7 str. Sakai DNA, complete genome

GenBank: BA000007.2

[FASTA](#) [Graphics](#)

Go to:

LOCUS BA000007 5498450 bp DNA circular BCT 18-JAN-2008

DEFINITION Escherichia coli O157:H7 str. Sakai DNA, complete genome.

ACCESSION BA000007 AP002550-AP002569

VERSION BA000007.2 GI:47118301

DBLINK Project: [226](#)

KEYWORDS .

SOURCE Escherichia coli O157:H7 str. Sakai

ORGANISM [Escherichia coli O157:H7 str. Sakai](#)

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.

REFERENCE 1

AUTHORS Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S., Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T., Yamamoto,K., Onishi,M., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and Shinagawa,H.

TITLE Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak

Genes Genet Syst 74 (5) 227-239 (1999)

Customize view

Basic Features

Gene, RNA, and CDS features only

Default features

Display options

Show sequence

Show reverse complement

Update View

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

LinkOut to external resources

Institute for Transcriptional Informatics [Institute for Transcriptional...]

REBASE enzyme XfaMrrP [REBASE - The Restriction Enzy...]

CC BY © 2012 中村保一(国立遺伝学研究所) licensed under CC表示2.1日本

豊富なリンクとツール群

論文

© 2012 中村保一 (国立遺伝学研究所) licensed under CC表示2.1日本

BLASTによる類似配列検索

特異的 primer デザイン

PubMed

NCBI

Escherichia coli O157:H7 str. Sakai DNA, complete genome

GenBank: BA000007.2

FASTA Graphics

Go to:

LOCUS BA000007 5498450 bp DNA circular BCT 18-JAN-2008

DEFINITION Escherichia coli O157:H7 str. Sakai DNA, complete genome.

ACCESSION BA000007 AP002550-AP002569

VERSION BA000007.2 GI:47118301

DBLINK Project: 226

KEYWORDS

SOURCE Escherichia coli O157:H7 str. Sakai

ORGANISM Escherichia coli O157:H7 str. Sakai

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.

REFERENCE 1

AUTHORS Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S., Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T., Yamamoto,K., Onishi,M., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and Shinagawa,H.

TITLE Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak

JOURNAL Genes Genet. Syst. 74 (5), 227-239 (1999)

PMID 10734605

REFERENCE 2

AUTHORS Ohnishi,M., Murata,T., Nakayama,K., Kiuchi,T., Kurokawa,K., Yasunaga,T., Yokoyama,K., Matsunaga,T., Hayashi,T.

TITLE Comparative analysis of the whole set of enterohemorrhagic Escherichia coli O157:H7 and Escherichia coli K-12 strain MG1655

JOURNAL Syst. Appl. Microbiol. 23 (3), 315-324

PMID 10734605

Display Settings: GenBank

Send: Change region shown

Customize view

Basic Features

Display options

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence

LinkOut to external databases

Institute for Transcriptional Regulation [Institute for Transcriptional Regulation]

REBASE enzyme XfaMrP [REBASE - The Restriction Enzyme Database]

REBASE enzyme M.Ecop933DamP [REBASE - The Restriction Enzyme Database]

REBASE enzyme S.EcoKO157ORFAP [REBASE - The Restriction Enzyme Database]

REBASE enzyme M.EphHK97DamP [REBASE - The Restriction Enzyme Database]

REBASE enzyme M.EcoCR63FP [REBASE - The Restriction Enzyme Database]

REBASE enzyme EcoKO157ORF5262P [REBASE - The Restriction Enzyme Database]

REBASE enzyme M.EcoKO157ORE1953P [REBASE - The Restriction Enzyme Database]

うーむ、がんばらねば

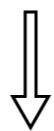
【New】

DB's and Services for

NGS

NGS

× Next-Generation Sequencer



○ New Generation Sequencer

example of the NGS's



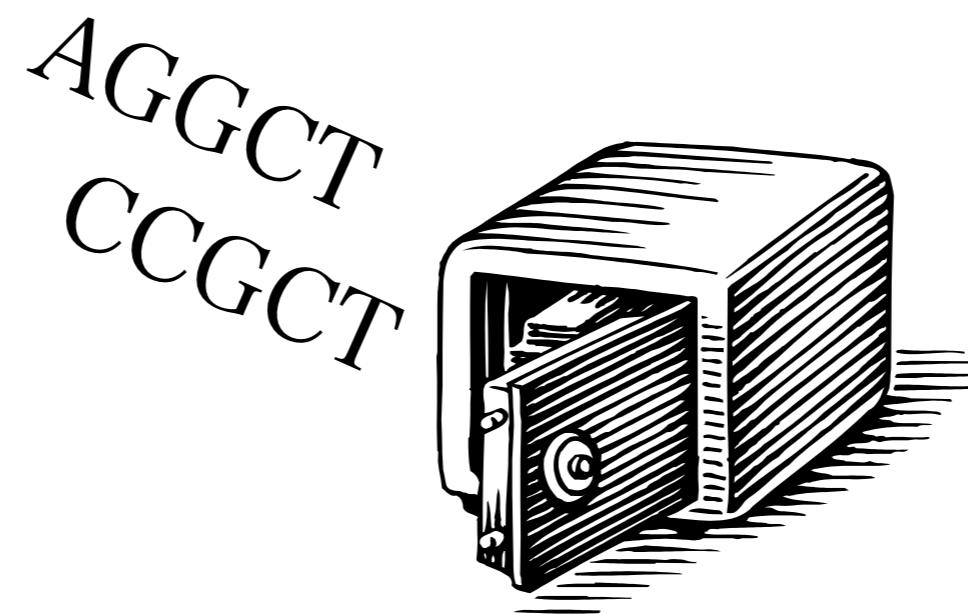
Roche (454): GS FLX+ System

illumina: Genome Analyzer IIx System

Life Technologies: 5500 xl SOLiD System

DDBJ Sequence Read Archive (DRA)

新型シーケンサデータの保管庫



DDBJ 配列リードアーカイブ (DRA)



DDBJ Sequence Read Archive

» Login D-way

DDBJ Sequence Read Archive

DDBJ Trace Archive

Home Documentation Submission Search Download Pipeline About

▶ English

DDBJ Sequence Read Archive (DRA) は Roche 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD® System などの次世代シークエンサからの出力データのためのデータベースです。DRA は International Nucleotide Sequence Database Collaboration (INSDC) のメンバーであり、NCBI Sequence Read Archive (SRA) と EBI Sequence Read Archive (ERA) との国際協力のもと、運営されています。従来のキャピラリ式シークエンサからの出力データは DDBJ Trace Archive にご登録ください。

[登録に必要なデータ](#)

[登録方法](#)

[データの検索・ダウンロード](#)

[DDBJ Read Annotation Pipeline でデータを解析](#)

[2011-01-06] かずさ DNA 研究所が登録したバイオディーゼルを生産するナンヨウアブラギリ (*Jatropha curcas*) の全ゲノムデータ (DRA000305, DRA000306) と cDNA データ (DRA000303, DRA000304) を公開 » かずさ DNA 研究所のプレスリリース

[2011-01-05] SRA Lite フォーマットデータの FTP 提供を開始しました。SRA フォーマットデータについては NCBI SRA Handbook を参照してください。

[2010-12-27] ライフサイエンス統合データベースセンターからデータ転送方法の動画マニュアルが公開されました。

DRA はライフサイエンス統合データベースプロジェクトの一部であり、科学技術振興機構のバイオインフォマティクス推進センターに支援されています。

Copyright©DNA Data Bank of Japan. All Rights Reserved.

Last modified: Jan. 28, 2011

<http://trace.ddbj.nig.ac.jp/dra/>

DRA: 新型シークエンサデータを保存・共有



DDBJ Sequence Read Archive

DDBJ Sequence Read Archive

DDBJ Trace Archive

Home Documentation Submission Search Download Pipeline About

DDBJ Sequence Read Archive (DRA) は Roche 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD® のためのデータベースです。DRA は International Nucleotide Sequence Database Collaboration (INSDC) のメンバーで、EBI Sequence Read Archive (ERA) との国際協力のもと、運営されています。従来のキャビラリ式シークエンサからの出力データ

登録に必要なデータ

登録方法

データの検索・ダウンロード

DDBJ Read Annotation Pipeline でデータを解析

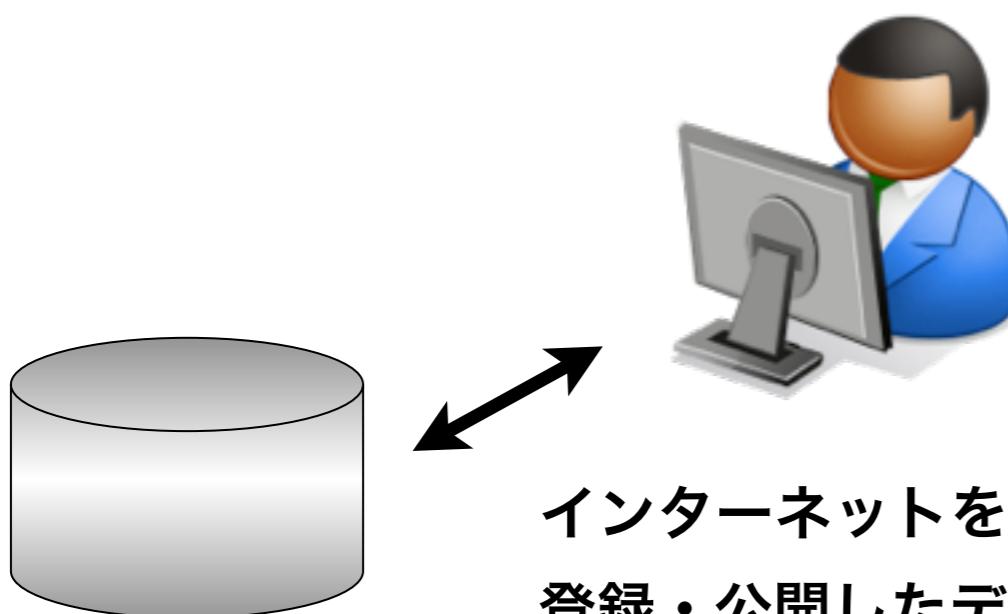
[2011-01-06] かずさ DNA 研究所が登録したバイオディーゼルを生産するナンヨウアブラギリ (*Jatropha curcas*) の全ゲノムデータ (DRA000303, DRA000304) を公開 ➤ かずさ DNA 研究所のプレスリリース

[2011-01-05] SRA Lite フォーマットデータの FTP 提供を開始しました。SRA フォーマットデータについては NCBI SRA Handbook

[2010-12-27] ライフサイエンス統合データベースセンターからデータ転送方法の動画マニュアルが公開されました。

DRA は ライフサイエンス統合データベースプロジェクトの一部であり、科学技術振興機構のバイオインフォマティクス推進センター

Copyright©DNA Data Bank of Japan. All Rights Reserved.



インターネットを通じて、
登録・公開したデータを、
別の研究者が再利用できる

Released Data					
search result : 30					
Accession	Study Title	Organism(s)	Center Name	Release Date	
DRA000001	Whole genome sequencing of <i>Bacillus subtilis</i> subsp. <i>natto</i> BEST195	<i>Bacillus subtilis</i> subsp. <i>natto</i>	KEIO	2010-03-26	<input type="button" value="Reset"/>
DRA000002	Whole genome resequencing of <i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 168	<i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 168	KEIO	2010-03-26	<input type="button" value="Reset"/>
DRA000010	Whole genome shotgun sequences of <i>Oryza sativa</i> japonica variety, Koshihikari	<i>Oryza sativa</i> Japonica Group	NIAS	2010-03-31	<input type="button" value="Reset"/>
DRA000030	Whole-genome DNA methylation analysis in human breast cancer cell lines using MeDIP-seq	<i>Homo sapiens</i>	KUGSIPS	2010-03-03	<input type="button" value="Reset"/>
DRA000039	genetic variation detected in 206 <i>klebsiella pneumoniae</i> plasmids	<i>Klebsiella pneumoniae</i>	WMC	2009-12-14	<input type="button" value="Reset"/>
DRA000067	<i>B. anthracis</i> BA103 genome analysis	<i>Bacillus anthracis</i>	NED	2010-04-22	<input type="button" value="Reset"/>
DRA000068	<i>B. anthracis</i> BA104 genome analysis	<i>Bacillus anthracis</i>	NED	2010-04-22	<input type="button" value="Reset"/>
DRA000099	Whole SNPs analysis of ciprofloxacin resistance among <i>B. anthracis</i> strains	<i>Bacillus anthracis</i>	NED	2010-04-22	<input type="button" value="Reset"/>
DRA000070	Whole SNPs analysis of ciprofloxacin resistance among <i>B. anthracis</i> strains	<i>Bacillus anthracis</i>	NED	2010-04-22	<input type="button" value="Reset"/>
DRA000155	CAGE analysis of whole adult brain and whole embryo rat transcriptome	<i>Rattus norvegicus</i>	RIKEN_OSC	2010-03-17	<input type="button" value="Reset"/>
DRA000169	Linking new promoters to functional transcripts in small samples with nanoCAGE and CAGEscan	<i>Homo Sapiens</i>	RIKEN_OSC	2010-06-08	<input type="button" value="Reset"/>
DRA000205	A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness	<i>Homo Sapiens</i>	RIKEN_OSC	2010-07-23	<input type="button" value="Reset"/>
DRA000220	Whole genome sequencing of <i>Oryzias latipes</i> H-9R	<i>Oryzias latipes</i>	KEIO-SM	2010-08-16	<input type="button" value="Reset"/>
SRA002052	Toxoplasma gondii transcript sequencing project	<i>Toxoplasma gondii</i>	UT-MGS	2009-07-01	<input type="button" value="Reset"/>
SRA002053	<i>Glossina morsitans</i> transcript sequencing project	<i>Glossina morsitans</i>	UT-MGS	2009-07-01	<input type="button" value="Reset"/>
SRA002054	<i>Glossina morsitans</i> transcript sequencing project	<i>Glossina morsitans</i>	UT-MGS	2009-06-25	<input type="button" value="Reset"/>
SRA002055	<i>Anopheles stephensi</i> transcript sequencing project	<i>Anopheles stephensi</i>	UT-MGS	2009-07-01	<input type="button" value="Reset"/>
SRA002056	<i>Cryptosporidium parvum</i> transcript sequencing project	<i>Cryptosporidium parvum</i>	UT-MGS	2009-07-01	<input type="button" value="Reset"/>
SRA002057	<i>Plasmodium yoelii</i> transcript sequencing project	<i>Plasmodium yoelii</i>	UT-MGS	2009-09-22	<input type="button" value="Reset"/>

研究者

FTP ディレクトリ /ddbji.database/dra/ / ftp.ddbj.nig.ac.jp

この FTP サイトはエクスプローラーでは表示するには、ページをクリックして、エクスプローラーで FTP サイトを開くをクリックしてください。

Welcome to DDBJ FTP Archive, running on ftp.ddbj.nig.ac.jp!

Please contact ddbj@ddbj.nig.ac.jp when you have any problem for getting access to this archive, downloading the data, and etc.

Termination of DDBJ-XXX output format.

Here is the announcement.

<http://www.ddbj.nig.ac.jp/whatis/2010/040415-a.html>

A new directory, "output", was made under "ddbj_database".
Now, all of recent raw and sequence data for JGI and EBI are included in the new "output" directory.

For details, please read the README.TXT in this directory.

Distributions of the latest DDBJ release and newly-derived/modified entries after that release can be retrieved at the following FTP servers.

DDBJ Fasta file
ftp://ftp.ddbj.nig.ac.jp/inf/ddbj_database/fasta/
ftp://ftp.ddbj.nig.ac.jp/inf/ddbj_database/fasta/

DDBJ Fasta (TPIK)
ftp://ftp.ddbj.nig.ac.jp/inf/ddbj_database/fasta/
ftp://ftp.ddbj.nig.ac.jp/inf/ddbj_database/fasta/

DDBJ XXX
ftp://ftp.ddbj.nig.ac.jp/inf/ddbj_database/xxx/
ftp://ftp.ddbj.nig.ac.jp/inf/ddbj_database/xxx/

It was last modified on Fri Feb 26 2010.

1階層上のディレクトリ

07/01/2010 01:07年総	ディレクトリ DRA000
07/01/2010 01:07年総	ディレクトリ DRA001
07/01/2010 01:08年総	ディレクトリ DRA002
07/01/2010 01:08年総	ディレクトリ DRA003
07/01/2010 01:09年総	ディレクトリ DRA004
07/01/2010 01:09年総	ディレクトリ DRA005
07/01/2010 01:10年総	ディレクトリ DRA006
07/01/2010 01:10年総	ディレクトリ DRA007
07/01/2010 01:11年総	ディレクトリ DRA008
07/01/2010 01:11年総	ディレクトリ DRA009
07/01/2010 01:12年総	ディレクトリ DRA010
07/01/2010 01:12年総	ディレクトリ DRA011
07/01/2010 01:13年総	ディレクトリ DRA012
07/01/2010 01:13年総	ディレクトリ DRA013
07/01/2010 01:14年総	ディレクトリ DRA014
07/01/2010 01:14年総	ディレクトリ DRA015
07/01/2010 01:15年総	ディレクトリ DRA016
07/01/2010 01:15年総	ディレクトリ DRA017
07/01/2010 01:16年総	ディレクトリ DRA018
07/01/2010 01:16年総	ディレクトリ DRA019
07/01/2010 01:17年総	ディレクトリ DRA020
07/01/2010 01:17年総	ディレクトリ DRA021
07/01/2010 01:18年総	ディレクトリ DRA022
07/01/2010 01:18年総	ディレクトリ DRA023
07/01/2010 01:19年総	ディレクトリ DRA024
07/01/2010 01:19年総	ディレクトリ DRA025
07/01/2010 01:20年総	ディレクトリ DRA026
07/01/2010 01:20年総	ディレクトリ DRA027
07/01/2010 01:21年総	ディレクトリ DRA028
07/01/2010 01:21年総	ディレクトリ DRA029
07/01/2010 01:22年総	ディレクトリ DRA030
07/01/2010 01:22年総	ディレクトリ DRA031
07/01/2010 01:23年総	ディレクトリ DRA032
07/01/2010 01:23年総	ディレクトリ DRA033
07/01/2010 01:24年総	ディレクトリ DRA034
07/01/2010 01:24年総	ディレクトリ DRA035
07/01/2010 01:25年総	ディレクトリ DRA036
07/01/2010 01:25年総	ディレクトリ DRA037
07/01/2010 01:26年総	ディレクトリ DRA038
07/01/2010 01:26年総	ディレクトリ DRA039
07/01/2010 01:27年総	ディレクトリ DRA040
07/01/2010 01:27年総	ディレクトリ DRA041
07/01/2010 01:28年総	ディレクトリ DRA042
07/01/2010 01:28年総	ディレクトリ DRA043
07/01/2010 01:29年総	ディレクトリ DRA044
07/01/2010 01:29年総	ディレクトリ DRA045
07/01/2010 01:30年総	ディレクトリ DRA046
07/01/2010 01:30年総	ディレクトリ DRA047
07/01/2010 01:31年総	ディレクトリ DRA048
07/01/2010 01:31年総	ディレクトリ DRA049
07/01/2010 01:32年総	ディレクトリ DRA050
07/01/2010 01:32年総	ディレクトリ DRA051
07/01/2010 01:33年総	ディレクトリ DRA052
07/01/2010 01:33年総	ディレクトリ DRA053
07/01/2010 01:34年総	ディレクトリ DRA054
07/01/2010 01:34年総	ディレクトリ DRA055
07/01/2010 01:35年総	ディレクトリ DRA056
07/01/2010 01:35年総	ディレクトリ DRA057
07/01/2010 01:36年総	ディレクトリ DRA058
07/01/2010 01:36年総	ディレクトリ DRA059
07/01/2010 01:37年総	ディレクトリ DRA060
07/01/2010 01:37年総	ディレクトリ DRA061
07/01/2010 01:38年総	ディレクトリ DRA062
07/01/2010 01:38年総	ディレクトリ DRA063
07/01/2010 01:39年総	ディレクトリ DRA064
07/01/2010 01:39年総	ディレクトリ DRA065
07/01/2010 01:40年総	ディレクトリ DRA066
07/01/2010 01:40年総	ディレクトリ DRA067
07/01/2010 01:41年総	ディレクトリ DRA068
07/01/2010 01:41年総	ディレクトリ DRA069
07/01/2010 01:42年総	ディレクトリ DRA070
07/01/2010 01:42年総	ディレクトリ DRA071
07/01/2010 01:43年総	ディレクトリ DRA072
07/01/2010 01:43年総	ディレクトリ DRA073
07/01/2010 01:44年総	ディレクトリ DRA074
07/01/2010 01:44年総	ディレクトリ DRA075
07/01/2010 01:45年総	ディレクトリ DRA076
07/01/2010 01:45年総	ディレクトリ DRA077
07/01/2010 01:46年総	ディレクトリ DRA078
07/01/2010 01:46年総	ディレクトリ DRA079
07/01/2010 01:47年総	ディレクトリ DRA080
07/01/2010 01:47年総	ディレクトリ DRA081
07/01/2010 01:48年総	ディレクトリ DRA082
07/01/2010 01:48年総	ディレクトリ DRA083
07/01/2010 01:49年総	ディレクトリ DRA084
07/01/2010 01:49年総	ディレクトリ DRA085
07/01/2010 01:50年総	ディレクトリ DRA086
07/01/2010 01:50年総	ディレクトリ DRA087
07/01/2010 01:51年総	ディレクトリ DRA088
07/01/2010 01:51年総	ディレクトリ DRA089
07/01/2010 01:52年総	ディレクトリ DRA090
07/01/2010 01:52年総	ディレクトリ DRA091
07/01/2010 01:53年総	ディレクトリ DRA092
07/01/2010 01:53年総	ディレクトリ DRA093
07/01/2010 01:54年総	ディレクトリ DRA094
07/01/2010 01:54年総	ディレクトリ DRA095
07/01/2010 01:55年総	ディレクトリ DRA096
07/01/2010 01:55年総	ディレクトリ DRA097
07/01/2010 01:56年総	ディレクトリ DRA098
07/01/2010 01:56年総	ディレクトリ DRA099
07/01/2010 01:57年総	ディレクトリ DRA100
07/01/2010 01:57年総	ディレクトリ DRA101
07/01/2010 01:58年総	ディレクトリ DRA102
07/01/2010 01:58年総	ディレクトリ DRA103
07/01/2010 01:59年総	ディレクトリ DRA104
07/01/2010 01:59年総	ディレクトリ DRA105
07/01/2010 01:60年総	ディレクトリ DRA106
07/01/2010 01:60年総	ディレクトリ DRA107
07/01/2010 01:61年総	ディレクトリ DRA108
07/01/2010 01:61年総	ディレクトリ DRA109
07/01/2010 01:62年総	ディレクトリ DRA110
07/01/2010 01:62年総	ディ

DRAsearch: 検索システム



DDBJ Sequence Read Archive

DDBJ Sequence Read Archive

» Login D-way

Home Documentation Submission Search Download Pipeline About

▶ English

DDBJ Sequence Read Archive (DRA) は Roche 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD® System などの次世代シークエンサからの出力データのためのデータベースです。DRA は International Nucleotide Sequence Database Collaboration (INSDC) のメンバーであり、NCBI Sequence Read Archive (SRA) と EBI Sequence Read Archive (ERA) との国際協力のもと、運営されています。従来のキャビラリ式シークエンサからの出力データは DDBJ Trace Archive にご登録ください。

登録に必要なデータ

登録方法

データの検索・ダウンロード

DDBJ Read Annotation Pipeline でデータを解析

[2011-01-06] かずさ DNA 研究所が登録したバイオディーゼルを生産するナンヨウアブラギリ (*Jatropha curcas*) の全ゲノムデータ (DRA000305, DRA000306) と cDNA データ (DRA000303, DRA000304) を公開 » かずさ DNA 研究所のプレスリリース

[2011-01-05] SRA Lite フォーマットデータの FTP 提供を開始しました。SRA フォーマットデータについては NCBI SRA Handbook を参照してください。

[2010-12-27] ライフサイエンス統合データベースセンターからデータ転送方法の動画マニュアルが公開されました。

DRA は ライフサイエンス統合データベースプロジェクトの一部であり、科学技術振興機構のバイオインフォマティクス推進センターに支援されています。

Copyright©DNA Data Bank of Japan. All Rights Reserved.

Last modified: Jan. 28, 2011

DRAsearch: Arabidopsis を選択

The screenshot shows the DRAsearch interface with the following details:

- Search Bar:** Shows the URL trace.ddbj.nig.ac.jp/DRASearch/.
- Organism Filter:** The input field contains "Ara", and the dropdown menu lists:
 - Arabidopsis arenosa
 - Arabidopsis lyrata
 - Arabidopsis thaliana
 - Arachis duranensis
 - Arachis hypogaea
- Study Type and Platform Filters:** Empty dropdown menus for "StudyType" and "Platform".
- Statistics:**
 - Released Entries:** Table showing the count of different entry types:

Type	Count
Submission	62023
Study	9728
Experiment	123324
Sample	256417
Run	393965
 - Organism:** Table showing the number of entries for different organisms:

#	Organism Name	Study
1	Homo sapiens	881
2	unidentified	877
3	Mus musculus	471
 - Study Type:** Table showing the number of entries for different study types:

#	Study Type	Study
1	Whole Genome Sequencing	4879
2	Transcriptome Analysis	1446
3	Metagenomics	1150
 - Center Name [All List]:** Table showing the number of entries for different centers:

#	Center Name	Study
1	JGI	1592
2	GEO	1376
3	JCVI	1310
- Information:** Data Last Update 2012-03-02, WebSite Last Update 2011-06-20.

DRAsearch: RNASeq を選択

The screenshot shows the DRAsearch interface on a web browser. The URL in the address bar is `trace.ddbj.nig.ac.jp/DRAsearch/query?organism=Ara&study_type=¢er_name=&platform=&show=20&sort=Study`. The page displays search fields for Accession, Organism (Arabidopsis thaliana), CenterName, Keyword, and a dropdown for StudyType. The StudyType dropdown is open, showing a list of study types, with 'RNASeq' circled in red.

Accession :

Organism : StudyType :

CenterName : Platform :

Keyword :

Show 20 records Sort by Study

Search Results (0 studies)

Copyright©DNA Data Bank of Japan. All Rights Reserved.

Study Types (部分表示):

- 16S pyrosequencing
- 16S rRNA tag sequencing
- Cancer Genomics
- copy number variation
- Deep Amplicon Sequencing
- environmental amplicon sequencing
- Epigenetics
- Exome Sequencing
- Forensic or Paleo-genomics
- Functional Genomics
- Gene Regulation Study
- genome partial sequencing
- Genome Variation Profiling
- Genomics
- Metagenomics
- Metatranscriptomics
- multi-isolate
- Other
- Paleo-genomics
- Paleo-genomics array capture
- pooled clone
- Pooled Clone Sequencing
- Population Genomics
- RAD Sequencing
- Resequencing
- PIP-CIP
- RNASeq** (highlighted with a red circle)
- small RNA
- sorted chromosome sequencing
- subtractive hybridization
- Synthetic Genomics
- Taq ssu rRNA gene hypervariable region taq sequencing

DRAsearch: Arabidopsis & RNASeq

統合データベース × DNA Data Bank × DDBJ Sequence × Result List - D × SRP006839 × SRX065809 × Murasaki Project ×

← → C trace.ddbj.nig.ac.jp/DRAsearch/query?organism=Arabidopsis+thaliana&study_type=RNASeq¢er_name=&platform=&show... ☆ 🔍

W DBCLS E P R S Papers Tech travel Dilgolet その他のブックマーク

DRAsearch Send Feedback  Search Home 

Accession :

Organism : StudyType :

CenterName : Platform :

Keyword :

Show 20 records Sort by Study

Search Results (10 studies) << < 1 / 1 Page > >>

#	STUDY	SUBMISSION	STUDY_TITLE	STUDY_TYPE	ORGANISM	BASES	SUBMITTED	CENTER_NAME
1	SRP006839	SRA037191	High-resolution profiling of small RNAs in the <i>Arabidopsis thaliana</i> root	RNASeq	Arabidopsis thaliana	9.2G	2011-05-24	Duke University
2	SRP007763	SRA044892	Genome-wide detection of context-sensitive alternative splicing in <i>Arabidopsis</i> RNASeq roots	RNASeq	Arabidopsis thaliana	1.3G	2011-08-11	Institute of Plant and Microbial Biology, Academia
3	SRP008348	SRA046111	GSE32318: 2-week-old <i>Arabidopsis</i> seedlings (Columbia ecotype)	RNASeq	Arabidopsis thaliana	3.4G	2011-09-23	GEO
4	SRP008822	SRA046998	Unexpected diversity of chloroplast non-coding RNAs as revealed by deep sequencing of the <i>Arabidopsis</i> transcriptome	RNASeq	Arabidopsis thaliana	9.8G	2011-10-12	Salk-E
5	SRP009340	SRA048085	GSE33713: Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos	RNASeq	Arabidopsis thaliana		2011-11-15	GEO
6	SRP009413	SRA048172	GSE33866: Deep-sequencing of total RNA from manually dissected wild-type <i>Arabidopsis</i> embryos at the 2-to-4 cell	RNASeq	Arabidopsis thaliana		2011-11-22	GEO

trace.ddbj.nig.ac.jp/DRAsearch/study?acc=SRP006839

DRAsearch: SRP006839

The screenshot shows a web browser window with the URL trace.ddbj.nig.ac.jp/DRASearch/study?acc=SRP006839. The page title is "SRP006839". The main content area displays "Study Detail" information:

Title	High-resolution profiling of small RNAs in the <i>Arabidopsis thaliana</i> root
Abstract	Small noncoding RNAs (ncRNAs) are key regulators of plant development through modulation of the processing, stability and translation of larger RNAs. In this project we present small RNA datasets produced from over 200 million Illumina sequencing reads covering all major cell types of the root (epi .. [more])
Description	Columbia-0 ecotype. Cell types were isolated using fluorescence activated cell sorting as described in Birnbaum et al. (2003), and developmental zones were hand dissected. More details in library construction protocols. Associated publication is Breakfield, Corcoran, et al. (2011) High-resolution .. [more]
Project ID	
Center Name	Duke University

SRA Links

Entrez Link	Pubmed : 21940835
-------------	-----------------------------------

The right sidebar contains a "Navigation" section with a list of SRA entries, each with a download icon, a file name, and FASTQ and SRALite links. The entry "Experiment SRX065809" is circled in red.

Submission	SRA037191	FTP	
Experiment	SRX065809	FASTQ	SRALite
	SRX065853	FASTQ	SRALite
	SRX065854	FASTQ	SRALite
	SRX065856	FASTQ	SRALite
	SRX065857	FASTQ	SRALite
	SRX065858	FASTQ	SRALite
	SRX065859	FASTQ	SRALite
	SRX065860	FASTQ	SRALite
Sample	SRS208957		
	SRS208958		
	SRS208959		
	SRS208960		
	SRS209324		
	SRS209326		
	SRS209328		
	SRS209329		

Copyright©DNA Data Bank of Japan. All Rights Reserved.

DRAsearch: SRX065809

The screenshot shows a web browser window for the DRAsearch experiment SRX065809. The URL in the address bar is trace.ddbj.nig.ac.jp/DRASearch/experiment?acc=SRX065809. The page includes a navigation bar with links to various DDBJ databases and tools, and a main content area divided into sections for Experiment Detail and Library Description, along with a Navigation sidebar.

Experiment Detail

Title	Longitudinal Sections small RNA sequencing
Design Description	
Organism	Arabidopsis thaliana

Library Description

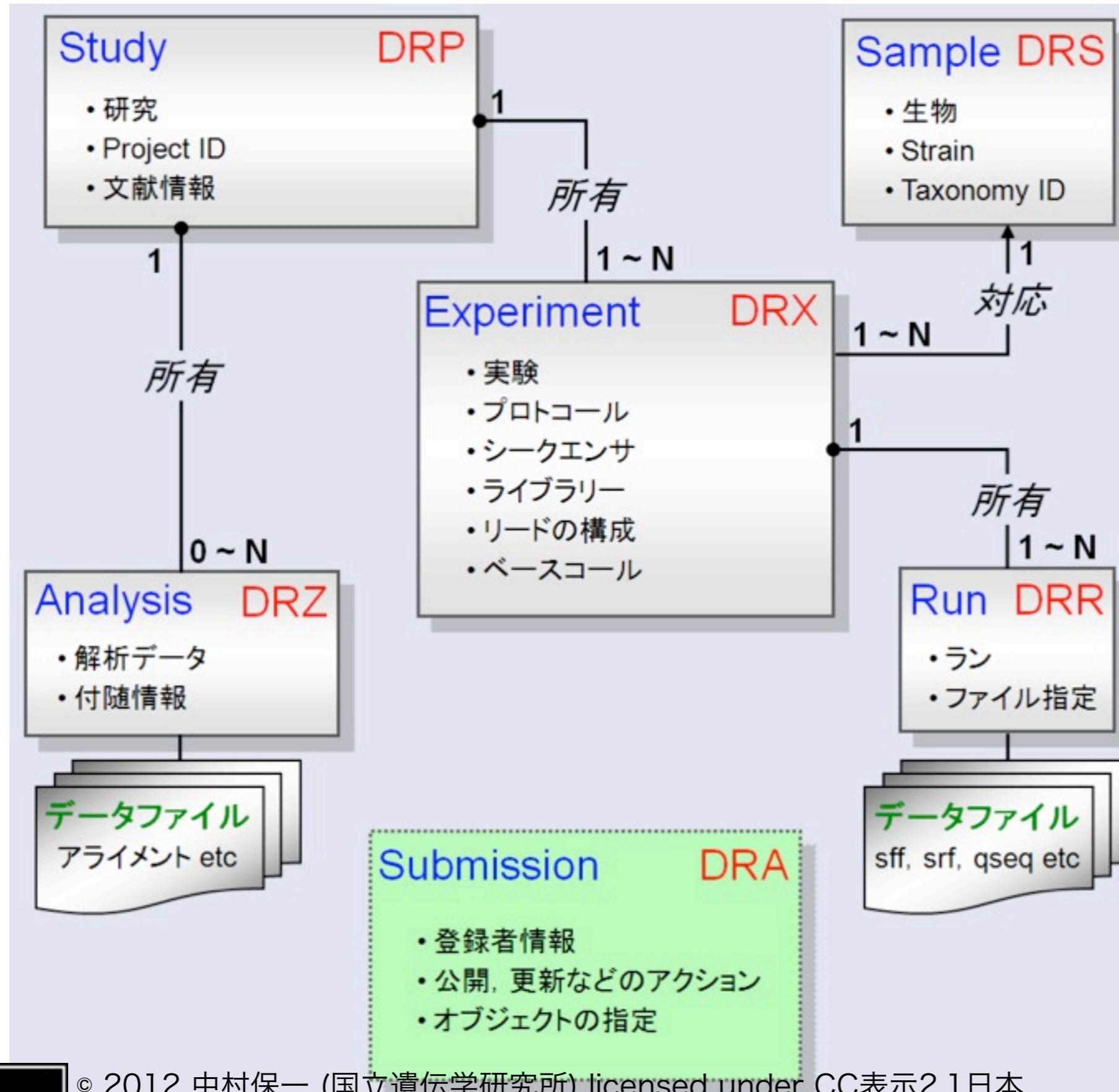
Name	LS1
Strategy	RNA-Seq
Source	TRANSCRIPTOMIC
Selection	size fractionation
Layout	SINGLE

Description: Cell type specific sorting was performed using GFP labeled lines (Birnbaum, Shasha et al. 2003). The stele was marked by pWOODEN LEG::GFP (WOL) (Mahonen, Bonke et al. 2000), endodermis and quiescent center by pSCARECROW::GFP (SCR) (Birnbaum, Shasha et al. 2003), the cortex by pCORTEX::GFP (COR) (Lee, Colinas et al. 2006), the epidermis and lateral root cap by pWEREWOLF::GFP (WER) (Lee and Schiefelbein 1999; Sena, Jung et al. 2004), and columella by enhancer trap PET111 (PET) (Nawy, Lee et al. 2005). At least 1 million GFP positive cells (or mock sorted cells in the case of whole root sorted samples) were collected directly into miRVana (Ambion) lysis buffer and stored at -80 degrees until extraction. The total RNA extraction protocol was used. For the longitudinal sections, 100 six-day-old Columbia-0 wild-type roots were hand dissected into 4 pieces: two meristematic zone

Navigation

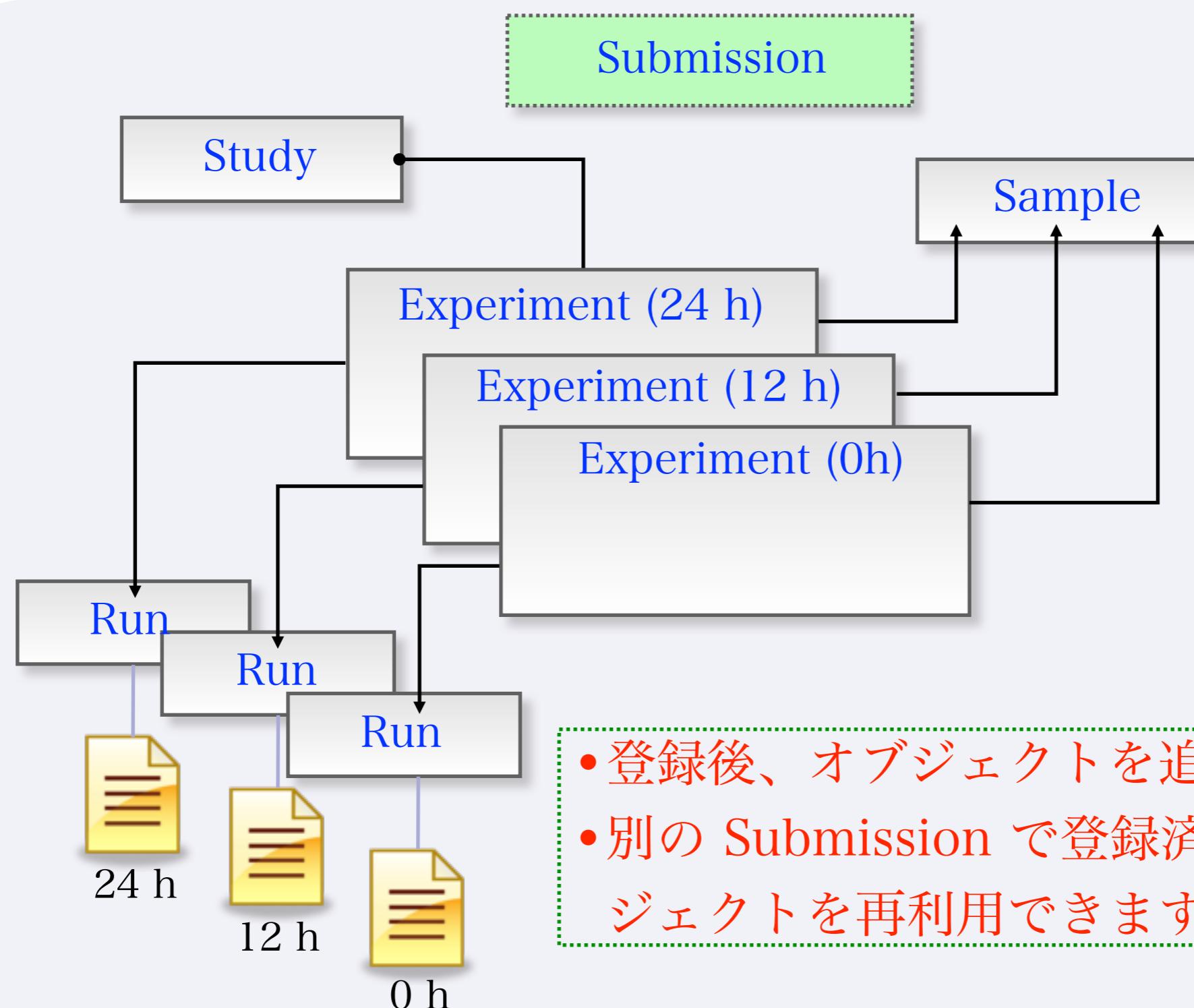
Submission	SRA037191	FTP	
Study	SRP006839		
Sample	SRS209324		
Run	SRR218085	FASTQ	SRALite
	SRR218086	FASTQ	SRALite
	SRR218087	FASTQ	SRALite
	SRR218088	FASTQ	SRALite

SRA の Metadata



メタデータ構成の例

例) 培養細胞: 薬剤処理 0, 12, 24 h 後の転写プロファイル解析



See: “Documentation”

The screenshot shows a web browser window for the DDBJ Sequence Read Archive. The URL in the address bar is trace.ddbj.nig.ac.jp/dra/documentation.shtml. The page title is "DDBJ Sequence Read Archive". A red circle highlights the "Documentation" link in the top navigation menu. Below the menu, there is a sidebar with links to "Home", "Documentation", "Submission", "Search", "Download", "Pipeline", and "About". The main content area starts with a section titled "1. 概要" (1. Overview) which describes the structure of metadata objects. At the bottom, there is a diagram illustrating the relationships between Study, DRP, and Sample DRS objects.

統合データ × DNA Data × DDBJ Seq. × Result Lis. × SRP00683 × DRA Met. × SRP00683 × SRX06580 × Murasaki ×

trace.ddbj.nig.ac.jp/dra/documentation.shtml

W DBCLS E P R S Papers Tech travel Dilgolet その他のブックマーク

DDBJ DNA Data Bank of Japan

DDBJ Sequence Read Archive DDBJ Trace Archive DDBJ BioProject

Home Documentation Submission Search Download Pipeline About » English

Home > Metadata

1. 概要

2. ガイドライン

3. XML スキーマ

4. メタデータの例

5. MetaDefine

1. 概要

メタデータにはランデータがどのようにして得られたのかが記載されています。メタデータは Submission, Study, Experiment, Sample, Run, Analysis の6つのオブジェクトから構成されます。各オブジェクトは XML スキーマで定義されており、オブジェクト同士は相互に関連付けられています。プレフィックスで区別されたアクセション番号が Submission (DRA), Study (DRP), Experiment (DRX), Sample (DRS), Run (DRR), Analysis (DRZ) オブジェクトに対して発行されます。メタデータとアクセション番号体系は DRA/ERA/SRA 間で共通です。DRA アクセション番号は論文中で引用することができます » 参考文献

« "メタデータオブジェクト" を非表示

Study DRP Sample DRS

• 研究
• Project ID

1

• 生物
• Strain

(cc) BY © 2012 中村保一 (国立遺伝学研究所) licensed under CC表示2.1日本

DDBJ の全「ゲノム」(WGS) 登録数の増大

year	count
2012.1-2	8
2011	37
2010	5
2009	4
2008	2
2007	(※13 Human Gut Metagenomes) 16
2006	5
2005	0
2004	1

DDBJ BioProject (New!)

Screenshot of the DDBJ BioProject Database homepage (www.ddbj.nig.ac.jp/index-j.html) in English.

The page features a navigation bar at the top with tabs for "Result List - DRA Search", "SRP006839 - DRA Search", and "SRX065809 - DRA Search". Below the navigation bar is a toolbar with links for "W", "DBCLS", "E", "P", "R", "S", "Papers", "Tech", "travel", and "Dligolet".

The main content area includes the DDBJ logo, a search bar, and language selection for ENGLISH. A sidebar on the left contains links for "HOME", "塩基配列の登録", "利用の手引き", "検索・解析", "FTP・WebAPI", "レポート・統計", and "お問い合わせ".

The central content area displays the title "DDBJ : DNA Data Bank of Japan" and a photograph of Mount Fuji. A sidebar on the right lists "Hot Topics" and "Maintenance" sections, both with a link to "一覧へ".

A red circle highlights the link "DDBJ BioProject Database" in the sidebar under "塩基配列の登録".

Hot Topics

- 2012.02.13 [DDBJ エントリへのリンク設定方法の変更](#)
- 2012.02.13 [DDBJ の新しいキーワード・エントリ検索システムについて](#)
- 2012.02.02 [ユーカリ \(*Eucalyptus camaldulensis*\) EST データの公開](#)

Maintenance

- 2012.02.20 [DDBJ サービスの中止・変更について \(2/15 現在の状況\)](#)
- 2012.02.14 (2/25) [国立遺伝学研究所ならびに DDBJ ネットワークの中止](#)
- 2012.02.13 [\(再開\) \(2012/2/23-27\) SAKURA によるデータ受付の一時休止](#)

Information

- [2012/3実施のコンピュータシステム移行に伴うお知らせ \(一覧\)](#)
- [DDBJ メールマガジン No.69 配信](#)

DDBJ BioProject とは

The screenshot shows a web browser window with the URL trace.ddbj.nig.ac.jp/bioproject/index.shtml. The page content describes BioProject as a database for research projects, managed by INSDC and DDBJ, and its connection to Sequence Read Archive and Trace Archive. It also mentions the PRJDB prefix for accessions. Below the main text, there's a section for 'DDBJ BioProject Search' and a sidebar with links to BioProject overview, registration requirements, and methods.

BioProject は研究プロジェクトとプロジェクトに由来するデータをまとめためのデータベースです。INSDC が運営するデータベースに登録されたデータが BioProject ID を引用することで、データがプロジェクト単位でグループ化されます。DDBJ では DDBJ 塩基配列データベース、Sequence Read Archive と Trace Archive にあるデータがまとめられます。BioProject はプロジェクトのゴール、実験材料や研究費の提供元といった情報を含んでいます。BioProject はゲノム配列決定プロジェクトを管理していた NCBI Genome Project を拡張し、デザインし直したものです。

DDBJ BioProject は登録されたプロジェクトに対して国際的に認可されたプレフィックス 'PRJD' のアクセス番号を発行します。公開されたプロジェクトデータは EBI/NCBI と交換されます。

DDBJ BioProject Search :現在 DDBJ からリリースされた PRJDB 番号のみを対象にしています

- » BioProject の概要
- » 登録に必要な情報
- » 登録方法

Diagram illustrating the relationship between DDBJ, BioProject, Sequence Read Archive, and Trace Archive:

```
graph TD; DDBJ((DDBJ)) <--> SP((BioProject)); SP <--> SRA((Sequence Read Archive)); SP <--> TA((Trace Archive))
```

The diagram shows four circular nodes: DDBJ at the top, BioProject in the center, Sequence Read Archive on the left, and Trace Archive on the right. Double-headed arrows connect BioProject to both SRA and TA, and a single-headed arrow points from DDBJ down to BioProject.

data

data

data

data

1
data

SRA 終了!?

NCBI Sequence Read Archive

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search SRA for Go Clear

About Entrez Related Resources Trace Archive SRA Trace Assembly

Limits Preview/Index History Clipboard Details

Sequence Read Archive (SRA) and Trace Archive repositories have been discontinued

Due to budget constraints, NCBI will be discontinuing its Sequence Read Archive (SRA) and Trace Archive repositories for high-throughput sequence data. Closure of the databases will occur in phases. SRA and Trace will stop accepting some types of submissions in the coming weeks, and all submissions within the next 12 months. Over the next several months, NCBI will be working with staff from NIH Institutes that fund large-scale sequencing efforts to develop an approach for future access to and storage of the existing data. NCBI will continue to support and develop information resources for biological data derived from next-generation sequencing such as genotypes, common variations, rare variations, sequence assemblies and gene expression data. We therefore encourage the research community to continue submissions of these data to the applicable databases, including:

1. RNA-Seq and epigenomic data to GEO
2. Variants, genotypes, phased haplotypes, and polymorphisms to dbVar, dbGaP and dbSNP
3. Genomic assemblies to GenBank/WGS
4. Transcript assemblies to GenBank/TSA
5. 16S ribosomal RNA and other targeted locus survey assemblies to GenBank

NCBI expects new applications will continue to emerge for next generation technology. We are excited to work with the community to develop strategies for archiving other summary experimental measures that are informative, efficient, and valuable to the biomedical research community.

For further information about submissions, contact [NCBI's Help Desk](#).

<http://www.ncbi.nlm.nih.gov/sra> (Feb. 2011)

EBI's response

EMBL-EBI will continue to support the Sequence Read Archive for raw data

Hinxton, 16 February 2011 - Because the rapid evolution of DNA sequencing technology has reduced the cost and time it takes to analyse genomes, an enormous amount of data is each day being submitted to, and retrieved from, the Sequence Read Archive (SRA), a public repository that preserves raw experimental DNA data.

Increasingly, journals and funding agencies are requiring that research leaders deposit their DNA data – for example from the 1000 Genomes Project – in the SRA, which is operated by the International Nucleotide Sequence Database Collaboration (INSDC). Until now, the SRA has been hosted and curated by INSDC partners the European Bioinformatics Institute (EMBL-EBI), the US National Center for Biotechnology Information (NCBI) and the DNA Databank of Japan (DDBJ). This week, NCBI announced that owing to budget constraints they will need to phase out submissions to the SRA over the next 8 to 12 months. EMBL-EBI will continue to support the archive, and to collaborate with NCBI on many other projects.

EMBL-EBI considers public, open-access nucleic acid sequence

disseminate next generation sequence data. As we face unprecedented data growth and ever-broadening applications for sequencing, this has become one of the most challenging areas of our work.

EMBL-EBI's approach to archiving sequence data depends on the deep involvement of the user community. We combine novel software compression techniques with judicious data reduction to achieve scalable storage of sequence data, and look to domain experts outside of our organisation to collaborate in developing the smoothest and most useful data submission and retrieval routes. The emerging ELIXIR framework for bioinformatics infrastructure in Europe provides an excellent forum to develop further this approach.

With community agreement on appropriate data reduction and the active community engagement to achieve smooth submission and retrieval pipelines, EMBL-EBI will continue to operate its SRA. But as all archives must justify their costs on the basis of future benefits, we will continue actively to consult and work with a broad community of scientists to understand the value of this and other life science archives. As a matter

http://www.ebi.ac.uk/ena/SRA_announcement_Feb_2011.pdf

Archive (ENA), the SRA helps disseminate data from large-scale studies, relieving those who generate the data from the

community in such a serious way would be communicated in good time to allow for orderly transitions.

EMBL-EBI will continue to work actively with NCBI on many

DDBJ's response

DDBJ will continue Sequence Raw Data Archiving

2011/2/22

Contents

- [1 DDBJ will continue Sequence Raw Data Archiving](#)
- [2 Present status of DDBJ Raw sequence data Archive](#)
 - [2.1 DDBJ Sequence Read Archive\(DRA\)\[2\]](#)
 - [2.1.1 Upon submission to DDBJ](#)
 - [2.1.2 The size of the DRA](#)
 - [2.2 Trace Archive at DDBJ \(DTA\)\[5\]](#)
 - [2.2.1 Upon submission to DDBJ](#)

DDBJ will continue Sequence Raw Data Archiving

DDBJ has been archiving raw data from Sanger sequencers and so called next-generation sequencers as a part of EBI/NCBI/DDBJ International Nucleotide Sequence Database Collaboration (INSDC), by receiving submissions from sequence centers mainly in Japan and also from several other countries.

The data submitted to DDBJ are processed into INSDC approved format and exchanged among NCBI and EBI frequently to make a same data set in either bank at the timing of data publication to minimize the inconvenience of data access from all over the world.

In light of the recent announcement that NCBI, who has been playing a hub in raw data archiving, will discontinue its Sequence Read Archive and Trace Archive repositories, DDBJ's archiving will be affected in the near future.

However, at this moment, DDBJ does not plan to discontinue either of the service to meet the demand of the domestic community as well as of the global one.

DDBJ has just started to formulate the plan to minimize the effect to the present activity of INSDC as well as the entire community in collaboration with other INSDC members.

<http://www.ddbj.nig.ac.jp/whatsnew/2011/DRA20110222.html>

NCBI will continue to operate the SRA



Status of the NCBI Sequence Read Archive (SRA)

Subsequent to an announcement in February 2011 that NCBI was planning to phase out the SRA due to funding constraints, NIH support has been provided that will enable the continuation of SRA. NCBI will continue to operate the SRA as NIH's primary archive of high-throughput sequencing data and as part of the international partnership of archives at the NCBI, the European Bioinformatics Institute and the DNA Database of Japan. Data submitted to any of the three organizations are shared among them.

The SRA is managing high-throughput sequencing data from many large studies funded by NIH Institutes. The SRA will also continue to archive high-throughput sequencing data that are associated with:

1. RNA-Seq, ChIP-Seq, and epigenomic data that are submitted to GEO
2. Genomic and transcriptomic assemblies that are submitted to GenBank
3. 16S ribosomal RNA data associated with metagenomics that are submitted to GenBank

It is NCBI's policy to make its publicly available data, including that in the SRA, available to others for redistribution so that they can provide value added services, such as tool sets for analyzing data and alternate interfaces. NCBI will continue work on new approaches for optimum storage and retrieval of raw sequencing data and their alignments.

For further information about these repositories, contact [NCBI's Help Desk](#).



<http://www.ncbi.nlm.nih.gov/About/news/13Oct2011.html>

コスト：ストレイジ/配列決定

1

10

将来... (コスト：ストレージ/配列決定)

1



1

さらに... (コスト：ストレージ/配列決定)

10



1

“NANOPORE” sequencer

<http://www.nanoporetech.com/technology/introduction-to-nanopore-sensing/introduction-to-nanopore-sensing>

The screenshot shows the Oxford Nanopore Technologies website. The header includes a search bar, contact us link, and navigation menu with Home, Technology (selected), About Us, News, and Careers. The main content area is titled "Introduction to nanopore sensing". It discusses the history of nanopore sensing, the company's intellectual property portfolio, nanopore fabrication, and sensing mechanisms. A diagram illustrates how different analytes (DNA, RNA, proteins) affect current flow through a nanopore.

Technology

- Introduction to nanopore sensing
- Biological nanopores
- Solid state nanopores
- The GridION system
- Single use cartridge
- GridION workflow
- GridION informatics
- MinION: a miniaturised sensing instrument
- Analytes and Applications: DNA, RNA, proteins
- Fields of use
- Publications

Introduction to nanopore sensing

The concept of using a nanopore as a biosensor was first proposed in the mid 1990s when nanopores were starting to be researched at academic institutions such as Oxford, Harvard and UCSC - all Oxford Nanopore collaborators. In an industrial setting, Oxford Nanopore was founded in 2005 to translate nanopore science into an electronics-based technology. The end-to-end system includes sample preparation, molecular analysis and informatics, and is designed to provide disruptive user benefits in a number of applications.

Oxford Nanopore has a broad [intellectual property](#) portfolio that includes internal innovation and collaborations with world leading nanopore researchers. This IP includes fundamental nanopore sensing techniques through to solid-state nanopore sensing technology including graphene.

Nanopore fabrication

A nanopore is, essentially, a nano-scale hole. This hole may be:

- [Biological](#): formed by a pore-forming protein in a membrane such as a lipid bilayer
- [Solid-state](#): formed in synthetic materials such as silicon nitride or graphene
- Hybrid: formed by a pore-forming protein set in synthetic material

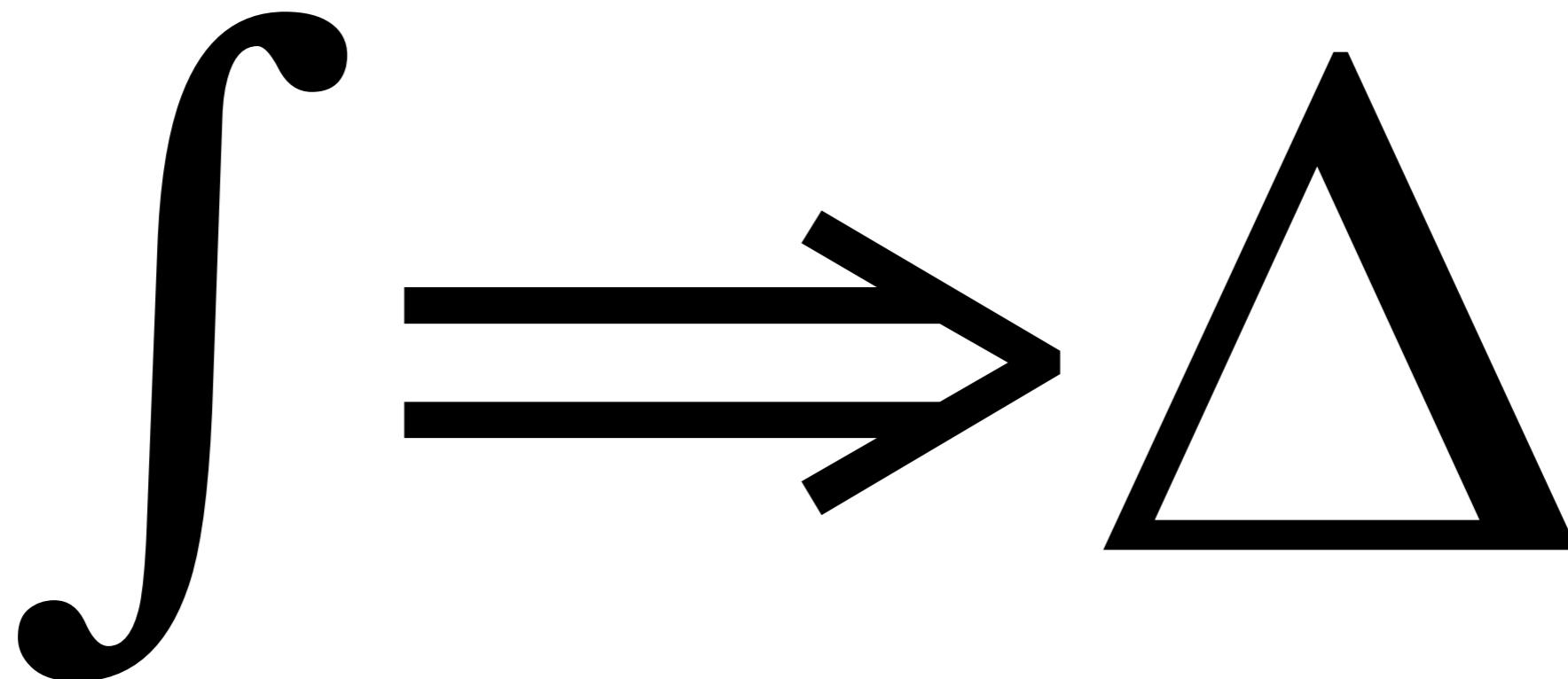
Nanopore sensing

A nanopore may be used to identify a target analyte as follows.

The diagram shows four stages of nanopore sensing:

- Two blue double helix molecules (DNA) are positioned above a central nanopore. Arrows labeled "Current flow" point downwards through the pore.
- A red circular molecule (RNA) is shown entering the nanopore from below, partially blocking it.
- A blue circular molecule (protein) is shown entering the nanopore from below, completely blocking it.
- A green oval-shaped molecule (large protein) is shown entering the nanopore from below, completely blocking it. A curved arrow points from the text "A nanopore may be used to identify a target analyte as follows." to this stage.

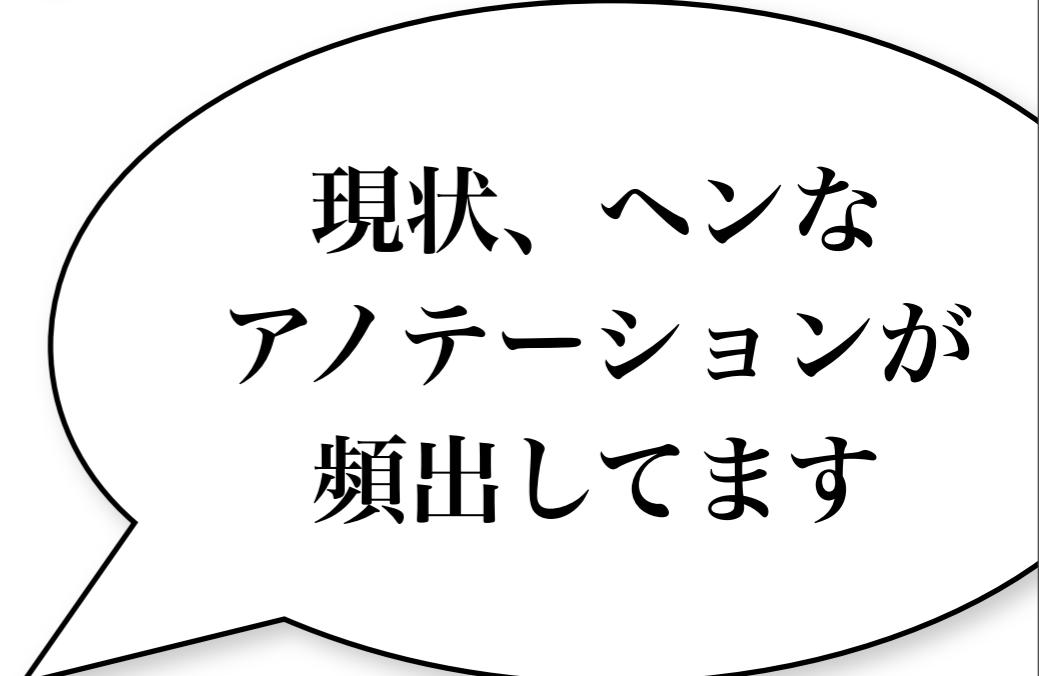
次々世代シーケンサに必要な配列DBとは



- ・ 良いリファレンス
- ・ 良い定型解析手順

Strange things in the Refseq / nr DB

- protain (183) < protein
- imilar to (28) < similar to
- simila to (22) < similar to
- cromosome (4) < chromosome
- RNA olymerase < polymerase
- dehydrogenas, ehydrogenase
- transposas, ransposase
- “2-Sep” for septin-2 < SEPT2



現状、ヘンな
アノテーションが
頻出しています

Identifier “mutation” by Excel (2-Sep)

The screenshot shows the BMC Bioinformatics journal website. At the top, there is a search bar with the placeholder "Search this journal for" and a "Go" button. To the left of the search bar is a yellow box displaying the "IMPACT FACTOR 3.03". The main navigation menu includes "Home", "Articles", "Authors", "Reviewers", "About this journal", and "My BMC Bioinformatics". On the far right, there is a link to "Advanced search".

On the left side, there is a sidebar with links to "Top", "Abstract", "Text", "Acknowledgements", and "References". Below this, a box titled "Learn to use resources in this article" contains a "GoMiner" link and a small DNA helix icon.

The main content area features a research article titled "Mistaken Identifiers: Gene name errors can be introduced inadvertently when using Excel in bioinformatics" by Barry R Zeeberg, Joseph Riss, David W Kane, Kimberly J Bussey, Edward Uchio, W Marston Linehan, J Carl Barrett, and John N Weinstein. The article is marked as "Highly accessed" and "Open access".

Below the title, it says "Highly accessed" and "Open access". The authors' names are listed: Barry R Zeeberg^{1†}, Joseph Riss^{2†}, David W Kane³, Kimberly J Bussey¹, Edward Uchio⁴, W Marston Linehan⁴, J Carl Barrett² and John N Weinstein^{1*}.

* Corresponding author: John N Weinstein weinstein@dtvpx2.ncifcrf.gov
† Equal contributors

Author Affiliations:

- 1 Genomics & Bioinformatics Group, Laboratory of Molecular Pharmacology, Center for Cancer Research (CCR), National Cancer Institute (NCI), National Institutes of Health (NIH), Bldg 37 Rm 5041, NIH, 9000 Rockville Pike, Bethesda, MD 20892 USA
- 2 Laboratory of Biosystems and Cancer, CCR, Bldg 37 Rm 5032, NIH, 9000 Rockville Pike, Bethesda, MD 20892 USA
- 3 SRA International, 4300 Fair Lakes CT, Fairfax, VA 22033 USA
- 4 Urologic Oncology Branch, Bldg 10 Rm 2B47, National Institutes of Health, Bethesda, MD 20892 USA

For all author emails, please [log on](#).

Associated material: PubMed record, About this article, Readers' comments (7).

Related literature: Cited by, Other articles by authors, on Google Scholar, on PubMed, Related articles/reviews.

BMC Bioinformatics 2004, 5:80 doi:10.1186/1471-2105-5-80

Identifier “mutation” made by Excel

The screenshot shows a Microsoft Excel spreadsheet titled "excel.gene2date.xls". The table consists of 15 rows and 11 columns. The columns are labeled A through K. The first column contains row numbers from 1 to 15. Columns A, B, C, E, F, H, I, J, and K contain gene names. Columns D and G are empty. The gene names are: APR-1, APR-2, APR-3, APR-4, APR-5, DEC-1, DEC-2, DEC1, DEC2, MAR1, MAR2, MAR3, NOV1, NOV2, and an empty cell at row 15. The "internal date format" is indicated by the text in the header cells above the table. The "default date format" is indicated by the date strings in the cells. The data is as follows:

	A	B	C	D	E	F	G	H	I	J	K
1	APR-1	35885	1-Apr		OCT-1	36068	1-Oct		SEP2	36039	2-Sep
2	APR-2	35886	2-Apr		OCT-2	36069	2-Oct		SEP3	36040	3-Sep
3	APR-3	35887	3-Apr		OCT-3	36070	3-Oct		SEP4	36041	4-Sep
4	APR-4	35888	4-Apr		OCT-4	36071	4-Oct		SEP5	36042	5-Sep
5	APR-5	35889	5-Apr		OCT-6	36073	6-Oct		SEP6	36043	6-Sep
6	DEC-1	36129	1-Dec		OCT1	36068	1-Oct		SEPT1	36038	1-Sep
7	DEC-2	36130	2-Dec		OCT11	36078	11-Oct		SEPT2	36039	2-Sep
8	DEC1	36129	1-Dec		OCT2	36069	2-Oct		SEPT3	36040	3-Sep
9	DEC2	36130	2-Dec		OCT3	36070	3-Oct		SEPT4	36041	4-Sep
10	MAR1	35854	1-Mar		OCT4	36071	4-Oct		SEPT5	36042	5-Sep
11	MAR2	35855	2-Mar		OCT6	36073	6-Oct		SEPT6	36043	6-Sep
12	MAR3	35856	3-Mar		OCT7	36074	7-Oct		SEPT7	36044	7-Sep
13	NOV1	36099	1-Nov		SEP-1	36038	1-Sep		SEPT8	36045	8-Sep
14	NOV2	36100	2-Nov		SEP-2	36039	2-Sep		SEPT9	36046	9-Sep
15					SEP1	36038	1-Sep				

BMC Bioinformatics 2004, 5:80 doi:10.1186/1471-2105-5-80

SEPT2 ⇒ 2-Sep case in Refseq

LOCUS XM_392412 2125 bp mRNA linear INV 12-APR-2011
DEFINITION PREDICTED: *Apis mellifera septin-2 (2-Sep)*, mRNA.
ACCESSION XM_392412
VERSION XM_392412.4 GI:328785636
KEYWORDS .
SOURCE *Apis mellifera* (honey bee)
ORGANISM *Apis mellifera*
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apidae; *Apis*.
COMMENT MODEL REFSEQ: This record is predicted by automated computational analysis. This record is derived from a genomic sequence (NW_003378075) annotated using gene prediction method: GNOMON supported by EST evidence.
Also see:
Documentation of NCBI's Annotation Process

On Apr 12, 2011 this sequence version replaced gi:110757583.

FEATURES Location/Qualifiers
source 1..2125
/organism="Apis mellifera"
/mol_type="mRNA"
/strain="DH4"
/db_xref="taxon:7460"
/linkage_group="LG6"
gene 1..2125
/gene="2-Sep"
/note="Derived by automated computational analysis using gene prediction method: GNOMON. Supporting evidence includes similarity to: 436 ESTs, 11 Proteins"
/db_xref="BEEBASE:GB17411"
/db_xref="GeneID:408882"
misc_feature 164..166
/gene="2-Sep"
/note="upstream in-frame stop codon"
CDS 194..1444
/gene="2-Sep"
/codon_start=1
/product="septin-2"
/protein_id="XP_392412.2"

septin-2 は
SEPT2
と記述される筈

http://www.ncbi.nlm.nih.gov/nuccore/XM_392412

How to Avoid such stupid Errors?

- よいリファレンス配列データセット
- よいリファレンスアノテーション
- 良い配列解析手順
- 遺伝子とメタデータのオントロジー
- 迷いとエラーの入らない配列解釈手順

- コピペ禁止
- エクセル禁止

MicrobeDB.jp project (JST/NBDC)

- Reference Microbe DBs & tools
 - Metagenome core DB & tools
 - Ortholog DB (MBGD)
 - Ontology (MEO)

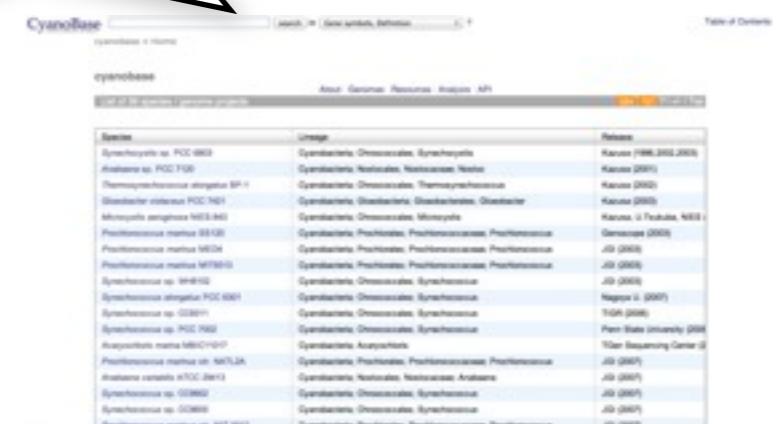
研生基
研山内

東工大（代表） 黒川研

KazusaAnnotation



TogoAnnotation



MicrobeDB.jp project のご紹介

連載 × DNA Data × DDBJ Seq. × Result Lis × SRP00683 × DRA Meta × SRP00683 × SRX06580 × Murasaki ×

events.biosciencedbc.jp/article

W DBCLS E P R S Papers Tech travel Dilgolet その他のブックマーク

NBDC National Bioscience Database Center NBDC の広報サイト バイオサイエンス ×DB=∞ 検索 Web events.biosciencedbc.jp/

Home シンポジウム 講習会 展示会 連載

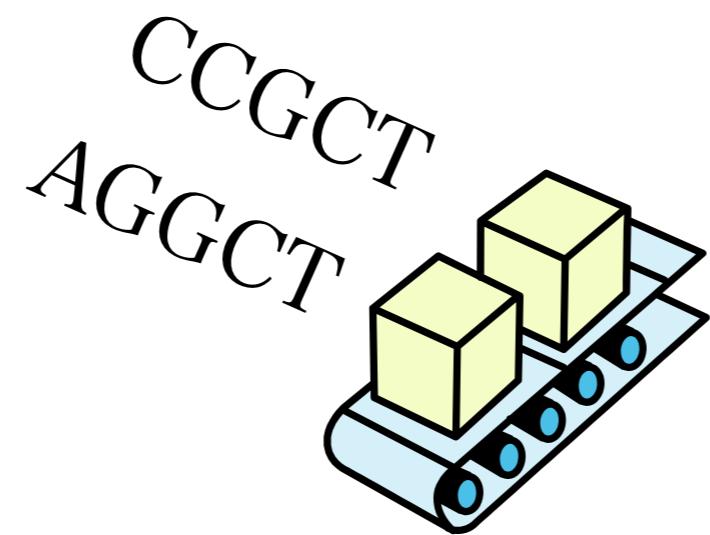
連載

『我が国のデータベース構築・統合戦略』（細胞工学連載の原稿を改変して掲載しています）

- 第1回 「データベースの現状と未来」
高木利久 バイオサイエンスデータベースセンター副センター長
- 第2回 「データベース統合利用基盤としてのセマンティックウェブ技術」
山口敦子 ライフサイエンス統合データベースセンター特任准教授
片山俊明 東京大学医科学研究所ヒトゲノム解析センター助教
- 第3回 「植物ゲノムデータベースの統合」
田畠哲之 かずきDNA研究所副所長
- 第4回 「微生物統合データベース「MicrobeDB.JP」」
黒川顕 東京工業大学大学院生命理工学研究科教授
- 第5回 「生体高分子の立体構造データベース（PDBj）とデータベース統合化」（2012年3月22日掲載予定）
中村春木 大阪大学蛋白質研究所附属プロテオミクス総合化学研究センター長・教授

DBJ 休止中

pipeline



DDBJ Pipeline: クラウド型解析ツール

連載 × DNA Data × DDBJ SeqL × Result Lis × SRP00683 × DRA Meta × SRP00683 × SRX06580 × Murasaki ×

www.ddbj.nig.ac.jp/index-j.html

W DBCLS E P R S Papers Tech travel Diigolet その他のブックマーク

HOME 塩基配列の登録 利用の手引き 検索・解析 FTP・WebAPI レポート・統計 お問い合わせ

▶ DDBJの紹介
▶ Q&A集
▶ 塩基配列の登録
SAKURA
大量登録システム(MSS)
データの修正・更新
DDBJ Sequence Read Archive
DDBJ Trace Archive
▶ プロジェクトの登録
DDBJ BioProject Database
▶ 検索
getentry
ARSA
TXSearch
BLAST
▶ 系統解析
ClustalW
▶ NGSデータ解析
DDBJ Read Annotation Pipeline
▶ ゲノム解析
GIB (GIB-GIB-V-GTPS)

DDBJ : DNA Data Bank of Japan

DDBJ（日本DNAデータバンク）は欧州と米国の対応機関（EBIおよびNCBI）と密接に協力しながら DDBJ/EMBL/GenBank 国際塩基配列データベースを構築している三大国際DNAデータバンクのひとつです

 Photo by Tatsuki Nagasaka

Hot Topics [一覧へ](#)

- 2012.02.13 DDBJ エントリへのリンク設定方法の変更
- 2012.02.13 DDBJ の新しい キーワード・エントリ検索システムについて
- 2012.02.02 ユーカリ (*Eucalyptus camaldulensis*) EST データの公開

Maintenance [一覧へ](#)

- 2012.02.20 DDBJ サービスの中止・変更について (2/15 現在の状況)
- 2012.02.14 (2/25) 国立遺伝学研究所ならびに DDBJ ネットワークの中止
- 2012.02.13 (再開) (2012/2/23-27) SAKURA によるデータ受付の一時休止

Information [一覧へ](#)

- 2012/3実施のコンピュータシステム移行に伴うお知らせ (一覧)
- DDBJ メールマガジン No.69 配信

塩基配列の登録・更新

■ 塩基配列の登録
塩基配列の登録手順を御案内します。

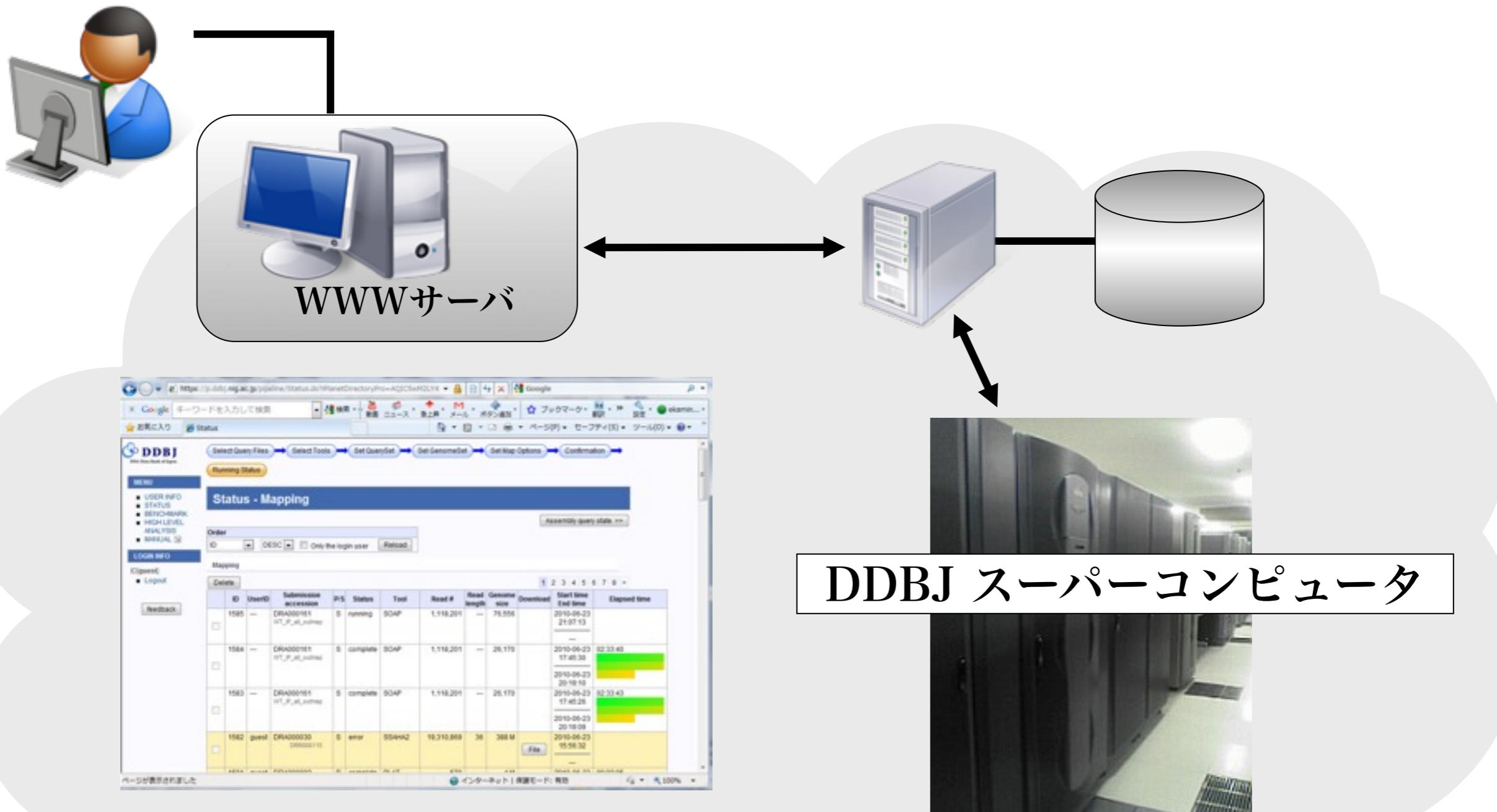
FTP・Web API

■ FTP (<ftp://ftp.ddbj.nig.ac.jp>)
DDBJリリースなどのデータファイルをダウンロード

 DDBJ
DNA Data Bank of Japan

DDBJ Pipeline: クラウド型解析ツール

研究者



DDBJ スーパーコンピュータ

計算資源が足りない → DDBJスパコンを使おう

解析できない → web ブラウザから自分で解析

DRA pipeline: ソフトウェア

よく用いられる
解析用ソフトウェアを
用意。クリックだけで
実行可能



ACCOUNT

login ID [guest]

[Logout]

ANALYSIS

Data setup

DRA Start

FTP Upload

HTTP Upload

step-1

Mapping / Assembly

step-2

Workflow

Genome (SNP/Short Indel)

RNA-seq (Tag count)

ChIP-seq

Genome (Large Indel)

Job Confirmation

step1-Mapping status

step1-Assembly status

step2-All status

Help

MANUAL

BENCHMARK

feedback



Selecting Tools for Basic Analysis of DDBJ ANNOTATION PIPELINE

BACK NEXT

Reference Genome Mapping

Tool	Help	Version	Input data			Evaluation			Analysis		Output format			Comment
			Base space	Color space	Paired end	Depth	Coverage	Error rate	SNP	Indel	.gff	.bed	SAM	
BLAT		34	✓			✓	✓	✓						Single-end analysis only
Mag		0.7.1	✓			✓	✓	✓	✓	✓	✓	✓	✓	
bwa		0.5.8a	✓			✓	✓	✓						✓
SSAHA2		2.3.0.1	✓			✓				✓	✓			SNP is single-end analysis only
SOAP		2.1.8	✓			✓	✓	✓	✓	✓	✓	✓		✓
Bowtie (SAMtools)		0.12.0 (0.1.7)	✓	✓		✓	✓	✓	✓	✓	✓			✓
TopHat		1.0.11 (BETA)	✓			✓	✓	✓						✓

de novo Assembly

Total limit = 22 Gbp

Tool	Help	Version	Base space	Color space	Paired-end	MSS(WGS)	Comment
Velvet		0.7.56	✓			✓	

DRA pipeline: 比較対象

イネ、マウスなど
解析比較対象となる

配列を多数用意

DDBJ
DNA Data Bank of Japan

ACCOUNT
login ID [guest]
[Logout](#)

ANALYSIS
step-1
[Mapping / Assembly](#)
step-2
[Genome \(SNP/Short Indel\)](#)
[Genome \(Large Indel\)](#)
[RNA-seq \(Tag count\)](#)
[ChIP-seq](#)
Job Confirmation
[step-1 Status](#)
[step-2 Status](#)
Help
[MANUAL](#)
[BENCHMARK](#)

Select Query Files → Select Tools → Set QuerySet → Set Genome

Running Status

Specifying Database of Reference Genome

Major genome sets

Organisms: **Oryza sativa japonica**

Genome sets:

- ✓ IRGSP Releases Build 4.0
- IRGSP Releases Build 5.0
- IRGSP Releases Build 5.0 masked by RepeatMasker with MIPS repeat data
- tigr version5.0
- tigr version6.0
- tigr version6.1
- tigr mitochondrial
- tigr chloroplast

User original sets

Organisms: **Mus musculus**

Genome sets:

- ✓ Jul. 2007 (mm9)
- Mar. 2006 (mm8)
- Aug. 2005 (mm7)
- NCBI build 36
- NCBI build 37

Organisms: **Arabidopsis thaliana**

Genome sets:

- ✓ TAIR8
- TAIR9

all check all clear

feedback

(cc) BY © 2012 中村保一 (国立遺伝学研究所) licensed under CC表示2.1日本

ネットワークを介してDDBJを使い倒す

遺伝研DDBJスーパーコンピュータ



DRA + DRA pipeline: 課題 = 非力

- DRA (DDBJ)

Strage: NFS 170 TB

- DRA pipeline (Mapping): an200 (DDBJ)

2x4 cores, 3.16GHz, 16GB memory x 33 nodes

Strage: NFS 105 TB

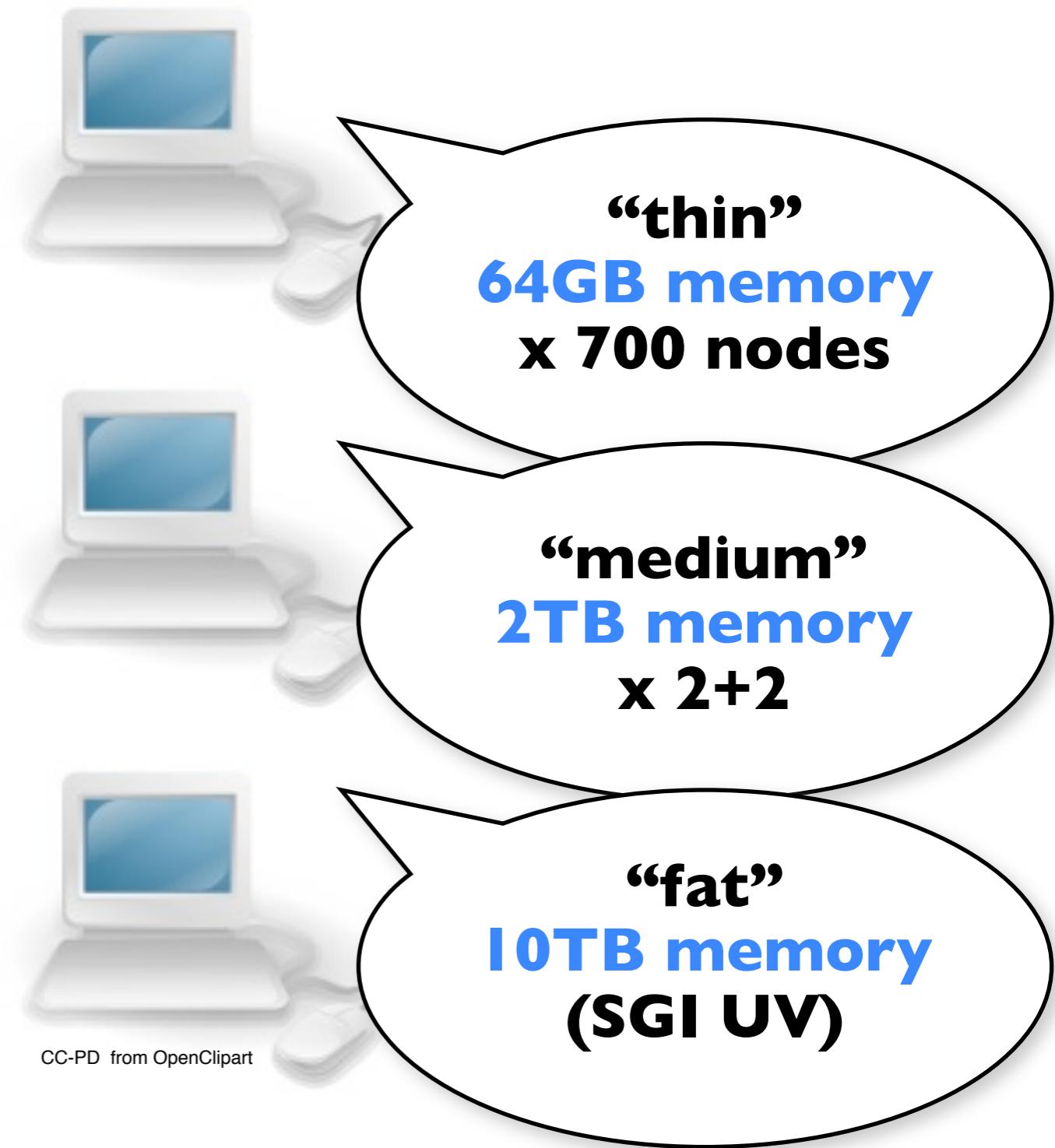
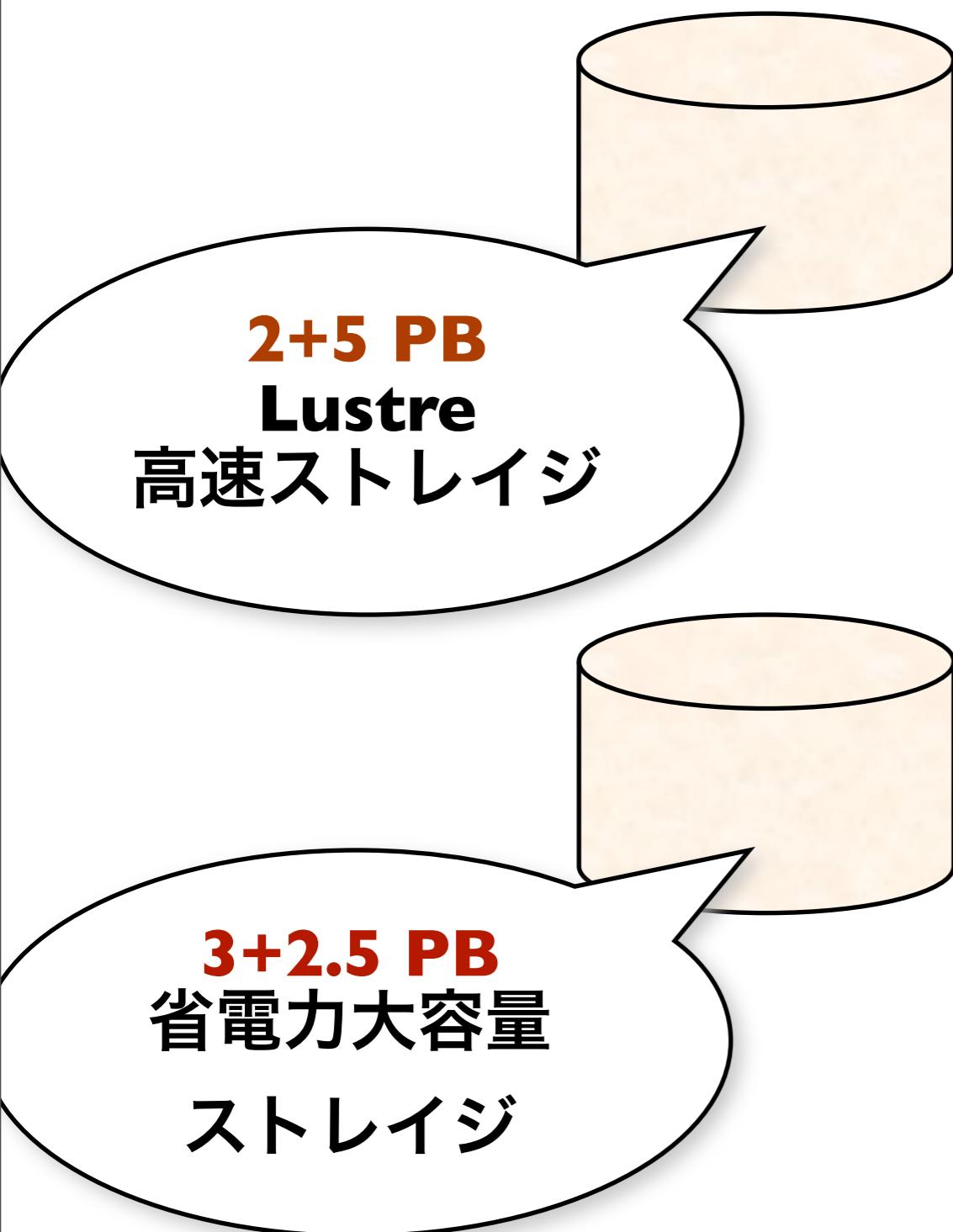
- DRA pipeline (Assemble): assembly1

4 cores, 2.49GHz, 256GB memory

- 開発用: Genome Informatics Laboratory's Takeru

4 nodes / 24 cores in total (24GB memory each)

新スーパーコンピュータ (from Mar. 2012)



未来の一つの形かもしれないかもかもしれない

DNA DropBox of Japan



Cloud strage with automatic pipeline

```
@SRR001654.1 9460:7:1:830:763 length=36
GTCATATTAAATCATACCAATACTCAAAAAATAA
+SRR001654.1 9460:7:1:830:763 length=36
I+-&*4)%+5#%)/&$%$##%"#&%%"$%#%"!%
@SRR001654.2 9460:7:1:402:781 length=36
GGCTAAAAAGCAAAATTCAAGTCTTCAAATAATT
+SRR001654.2 9460:7:1:402:781 length=36
II+($$+$%&*+/*("%"&+*%"&(*%"#%"&$
@SRR001654.3 9460:7:1:433:775 length=36
GTGCTTTTTTTCCAGGAAGTTGTCTCCTCTATC
+SRR001654.3 9460:7:1:433:775 length=36
II3DI>IIIIIB7.,&%&'&)."++,$"&$&"#%
```



low-data

- assemble

- mapping

- BLAST

- gene-finding



Report!

Acknowledgements

DRA, BioProject & DDBJ pipeline

are supported by Integrated Lifescience Database Project, BIRD and DDBJ.

Dr. Yuichi Kodama
(DDBJ DRA/DTA annotator)

Dr. Eli Kaminuma
(Assist. Prof. at Genome Informatics lab)

Dr. Hideki Nagasaki
(KAKENHI “genome support” Research Scientist)

Ms. Takako Mochizuki
(TRIC Research Scientist)

the other my lab’s members and
DDBJ people

TogoAnnotation & MicrobeDB.jp

is supported by Integrated Lifescience Database Project.

Prof. Hideaki Sugawara

Mr. Takatomo Fujisawa
(Project Research Scientist)

Dr. Shinobu Okamoto
(Project Assoc. Prof. at DBCLS)

the other my lab’s members,
MicrobeDB.jp project members
and the Kazusa people

ご清聴
ありがとうございました
にや

