

統合データベース講習会：AJACS駿河

2013年1月12日

DRA&DDBJパイプライン

情報・システム研究機構 (ROIS)
国立遺伝学研究所 DDBJセンター

中村 保一

NGS

New Generation Sequencer

⇒ データ爆発

代表的な新型シーケンサー



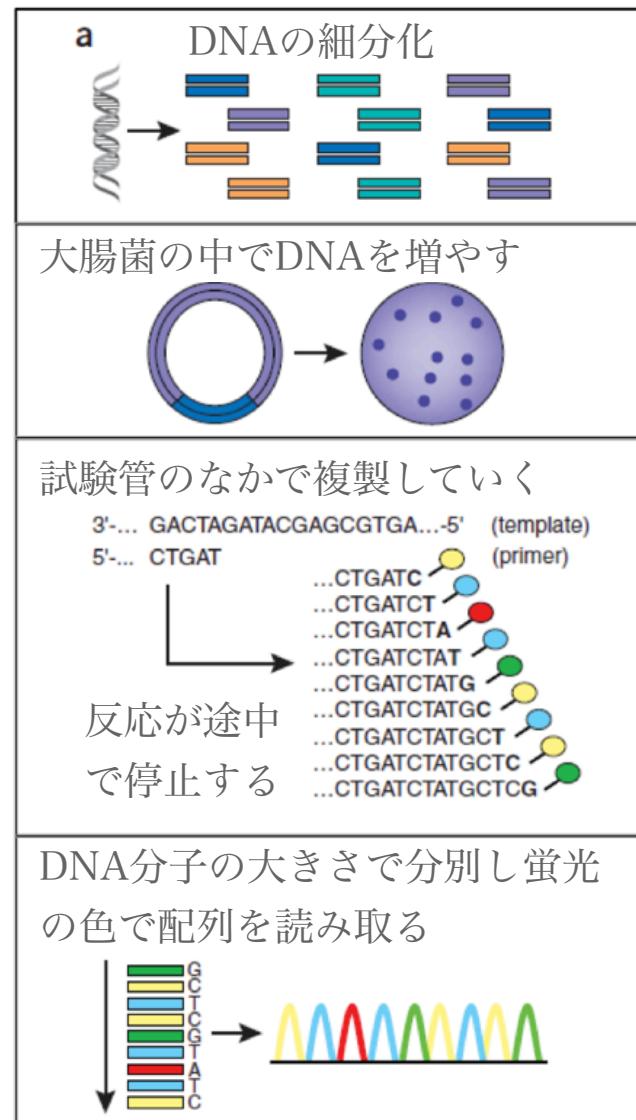
Roche (454): GS FLX+ System

illumina: Genome Analyzer IIx System

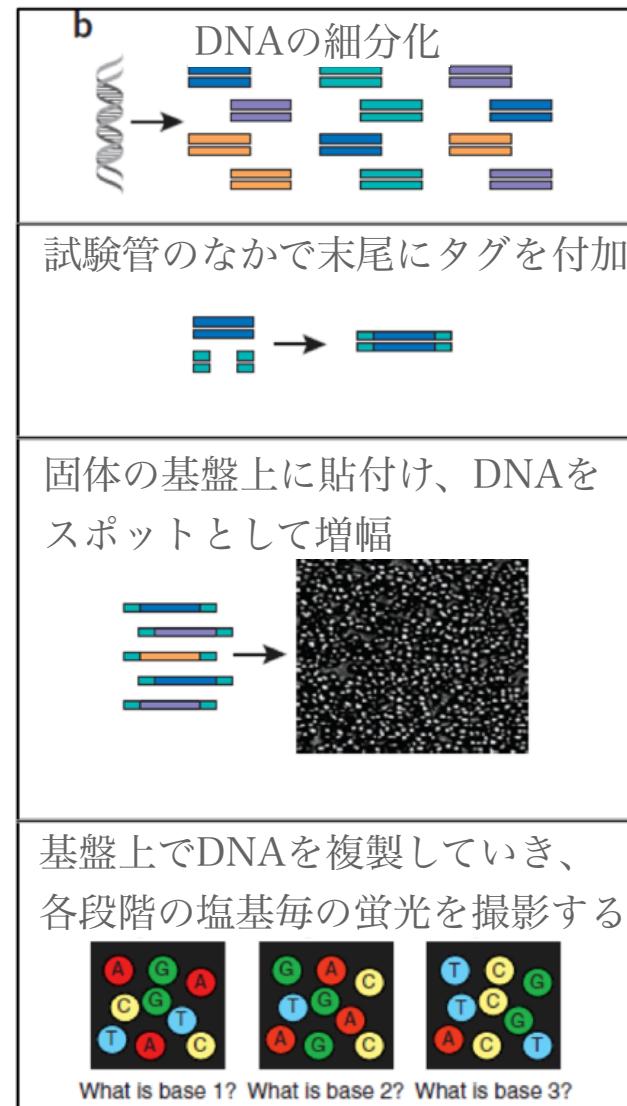
Life Technologies: 5500 xl SOLiD System

従来のシーケンサーと新型シーケンサー

従来法



新型

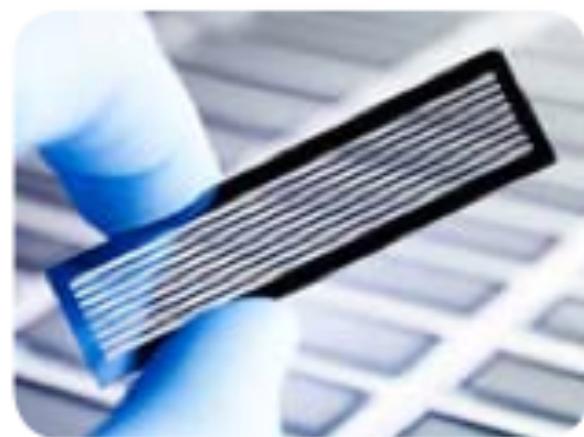


新型シークエンサはなぜ高速？→「集積度」

- 従来法は溶液やゲル中での反応と分離
- 固体担体を用いて超高密度化を可能にした



マイクロプレート
 $24 \times 16 = 384$ 穴



イルミナ社 GA フローセル
数千万スポット

NGSの例: illumina: GA の原理

<http://www.youtube.com/watch?v=77r5p8lBwJk>

- ⌚ フラットな固層上に適当な間隔でDNAを1分子ずつ固定、基盤上で「ブリッジPCR」を行い、スポットとしてDNAを増幅
- ⌚ 相補鎖合成を行いながら化学発光をとらえる
- ⌚ 4つの塩基に別々の蛍光標識をつけておいて、結合した塩基の場所をスポットの光として特定し、塩基配列を解読していく
- ⌚ 元データは時系列の高密度な画像データ

新型シークンサーの長所：高速・大量

	HiSeq 2000	HiSeq 1000
Output (2 × 100 bp)	600 Gb	300 Gb
Run Time (2 × 100 bp)	~11 days	~8.5 days
Cluster Generation	cBot	cBot
Paired-end Reads	6 Billion	3 Billion
Single Reads	3 Billion	1.5 Billion
Maximum Read Length**	2 × 100 bp	2 × 100 bp
Quality Scores***	> 85% (2 × 50 bp) > 80% (2 × 100 bp)	

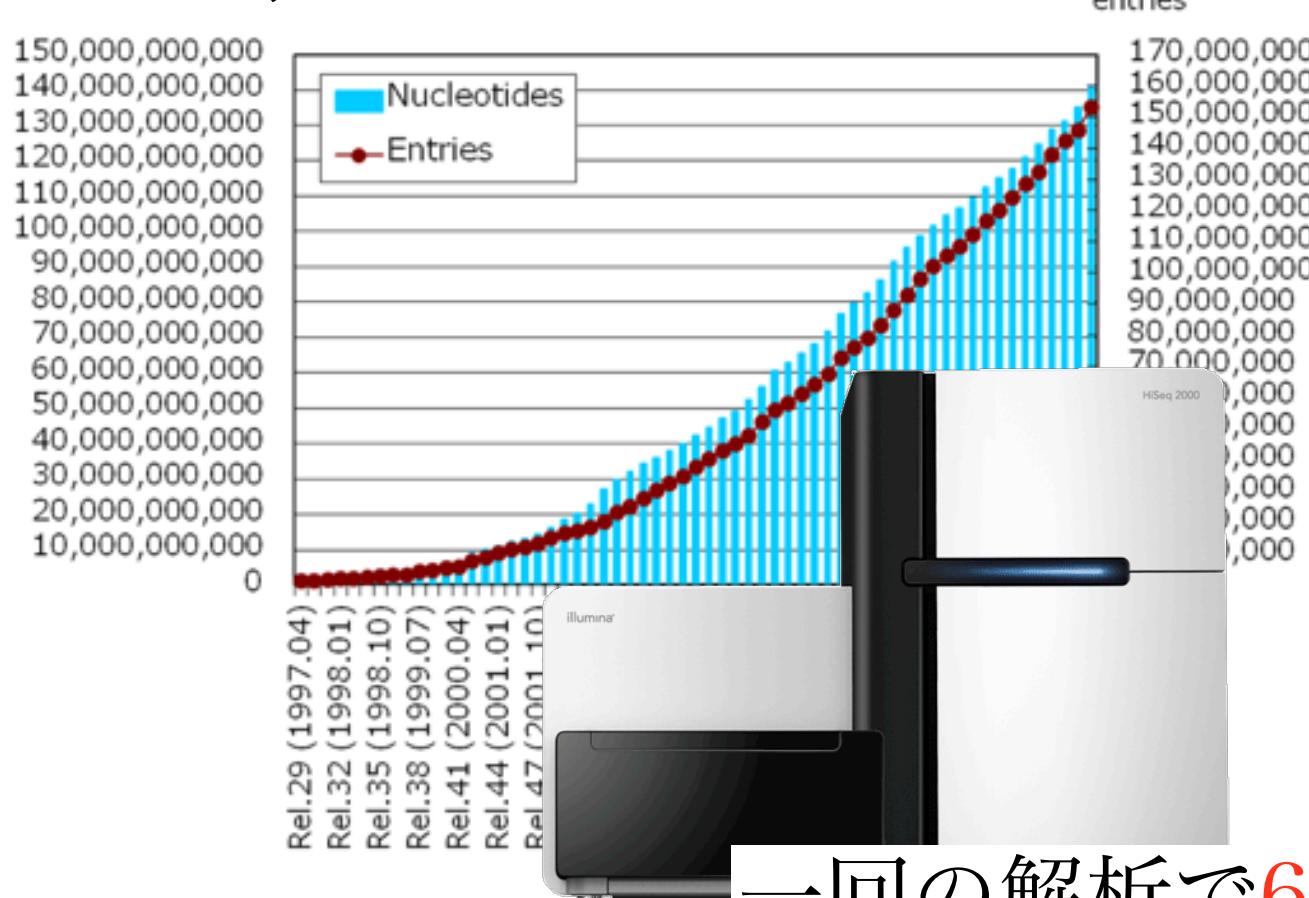


- イルミナの HiSeq2000
- 一解析で6000億塩基 (600ギガベース)
- ヒト一人のDNAがおよそ30億塩基対なので
- 一解析で200人前のデータが取得できる

問題点：データの爆発

DDBJ/EMBL/GenBank database growth

登録塩基数: 1,400億 = 140ギガ



一回の解析で600ギガ

“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 features the company logo and navigation links for Home, Technology, 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 nanopore sensing. A diagram illustrates current flow through a nanopore, and a photograph shows a hand holding the MinION sequencing device next to a laptop.

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 glass
- Hybrid: formed by a pore-forming protein set in synthetic material

Nanopore sensing

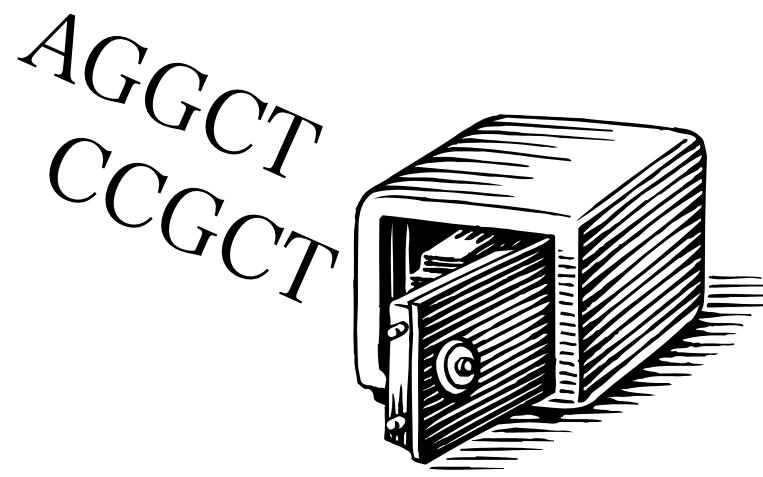
A nanopore may be used to identify a target analyte as follows.

Current flow

MinION - \$900 usb-powered DNA sequencer

新型シーケンサーデータ保管庫 (DRA)

新型シーケンサーからのデータをDDBJが保管することで、ユーザがデータを保管するために必要となる莫大なハードディスクの容量を肩代わりします！
そのかわり、データを公開して良い時期になったら公開し、共有・再利用できるように、提供します。



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

DDBJ BioProject

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 などの次世代シーケ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ケンサからの出力データは DDBJ Trace Archive にご登録ください。

登録に必要なメタデータとデータファイル

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

» 動画マニュアル

MetaDefine 動画マニュアル

データ転送動画マニュアル

[Site Policy](#) | [Privacy](#) | © DNA Data Bank of Japan

Last modified: Jan. 10, 2012

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

DRA: 新型シーケンサデータを保存・共有



DDBJ Sequence Read Archive

DDBJ Sequence Read Archive DDBJ Trace Archive DDBJ BioProject
Home Documentation Submission Search Download Pipeline About

DDBJ Sequence Read Archive (DRA) は Roche 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD® プラットフォーム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) の一員で、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。従来のシーケンサからの出力データは DDBJ Trace Archive にご登録ください。

登録に必要なメタデータとデータファイル

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

» 動画マニュアル

MetaDefine 動画マニュアル

データ転送動画マニュアル

Site Policy | Privacy | © DNA Data Bank of Japan

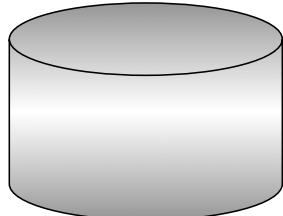
Released Data

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
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
DRA000010	Whole genome shotgun sequences of <i>Oryza sativa</i> japonica variety, Koshihikari	<i>Oryza sativa</i> Japonica Group	NIAS	2010-03-31
DRA000030	Whole-genome DNA methylation analysis in human breast cancer cell lines using MeDIP-seq	<i>Homo sapiens</i>	KUGSPS	2010-03-01
DRA000039	genetic variation detected in 206 <i>klebsiella pneumoniae</i> plasmids	<i>Klebsiella pneumoniae</i>	WMC	2009-12-14
DRA000067	<i>B. anthracis</i> BA103 genome analysis	<i>Bacillus anthracis</i>	NIID	2010-04-22
DRA000068	<i>B. anthracis</i> BA104 genome analysis	<i>Bacillus anthracis</i>	NIID	2010-04-22
DRA000069	Whole SNPs analysis of ciprofloxacin resistance among <i>B. anthracis</i> strains	<i>Bacillus anthracis</i>	NIID	2010-04-22
DRA000070	Whole SNPs analysis of ciprofloxacin resistance among <i>B. anthracis</i> strains	<i>Bacillus anthracis</i>	NIID	2010-04-22
DRA000155	CAGE analysis of whole adult brain and whole embryo rat transcriptome	<i>Rattus norvegicus</i>	RIKEN_OSC	2010-03-17
DRA000169	Linking new promoters to functional transcripts in small samples with nanoCAGE and CAGEscan	<i>Homo Sapiens</i>	RIKEN_OSC	2010-06-08
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
DRA000220	Whole genome sequencing of <i>Oryzias latipes</i> Hd-Rr	<i>Oryzias latipes</i>	KEIO-SM	2010-08-16
SRA002052	Toxoplasma gondii transcript sequencing project	<i>Toxoplasma gondii</i>	UT-MGS	2009-07-01
SRA002053	<i>Glossina morsitans</i> transcript sequencing project	<i>Glossina morsitans</i>	UT-MGS	2009-07-01
SRA002054	<i>Glossina morsitans</i> transcript sequencing project	<i>Glossina morsitans</i>	UT-MGS	2009-06-25
SRA002055	Anopheles stephensi transcript sequencing project	<i>Anopheles stephensi</i>	UT-MGS	2009-07-01
SRA002056	<i>Cryptosporidium parvum</i> transcript sequencing project	<i>Cryptosporidium parvum</i>	UT-MGS	2009-07-01
SRA002057	<i>Plasmodium yoelii</i> transcript sequencing project	<i>Plasmodium yoelii</i>	UT-MGS	2009-09-22

ページが表示されました

研究者



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

FTP ディレクトリ /ddbj_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 XML output format.
--Here is the announcement.
--http://www.ddbj.nig.ac.jp/whatsnew/2010/100825-e.html

--A new directory, "patent", was made under "ddbj_database".
--Now, all of patent amino acid sequence data for JPO and KIPO are
--Included in the new "patent" directory.

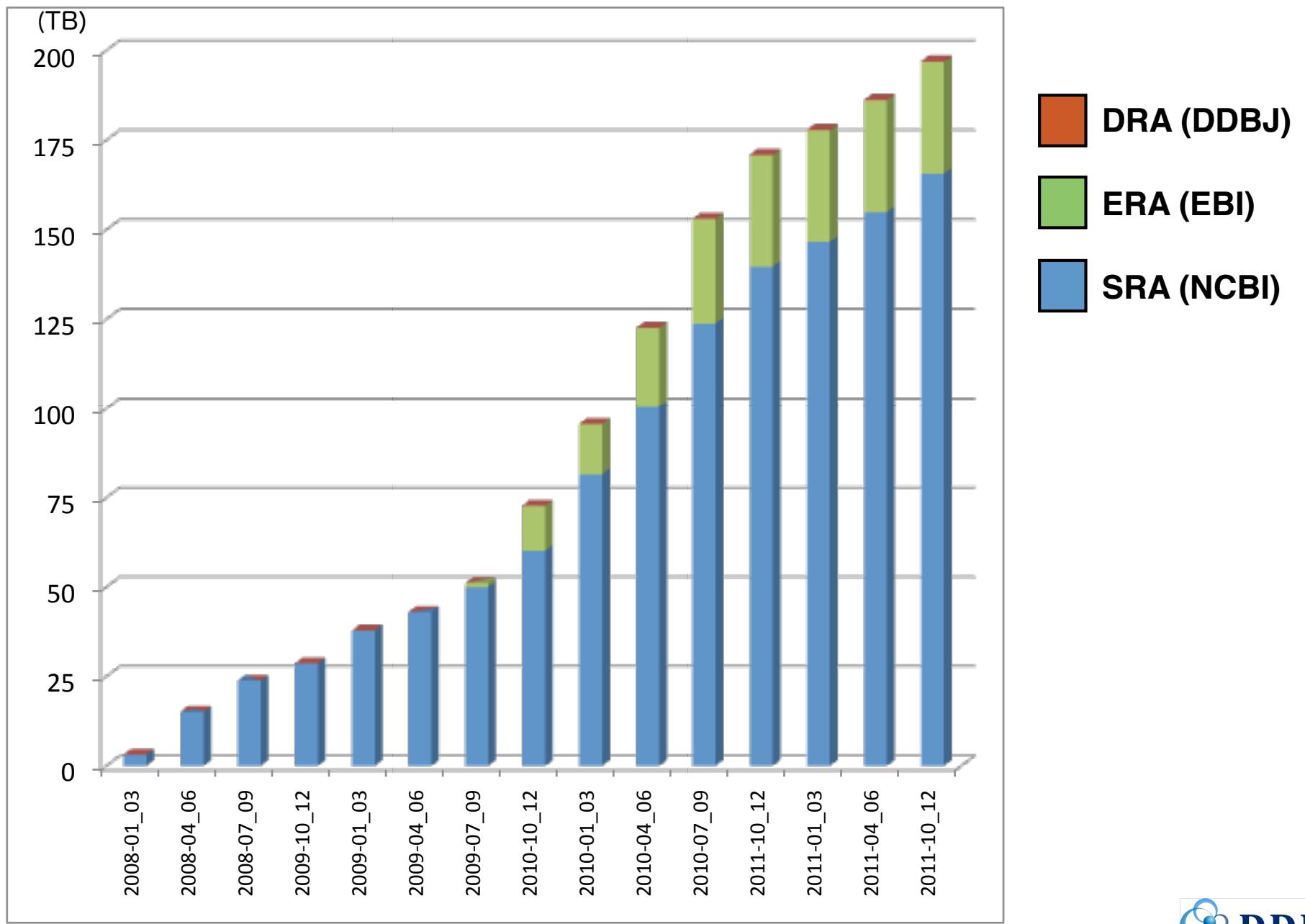
--For details, please read the README.TXT in this directory.

--Distribution of the latest DDBJ release and newly-arrived/updated
--entries after that release can be retrieved at the following FTP
--sites.
--DDBJ Flat file
--ftp://ftp.ddbj.nig.ac.jp/ddbj_database/ddbj/
--ftp://ftp.ddbj.nig.ac.jp/ddbj_database/ddbj/new/
--DDBJ Flat file(TPA)
--ftp://ftp.ddbj.nig.ac.jp/ddbj_database/tpo/ddbj/
--ftp://ftp.ddbj.nig.ac.jp/ddbj_database/tpo/ddbjnew/
--TPB XML
--ftp://ftp.ddbj.nig.ac.jp/ddbj_database/ddbj/xml/insdxml_current/
--ftp://ftp.ddbj.nig.ac.jp/ddbj_database/ddbjnew/xml/insdxml_current/
--It was last modified on Fri Feb 26 2010.
```

1階層上のディレクトリ

07/30/2010 08:07午後	ディレクトリ DRA000
07/05/2010 01:27午後	ディレクトリ ERA000
07/05/2010 01:27午後	ディレクトリ ERA001
07/30/2010 02:59午後	ディレクトリ ERA002
07/30/2010 08:03午後	ディレクトリ ERA003
07/30/2010 08:03午後	ディレクトリ ERA004
07/13/2010 04:54午後	ディレクトリ ERA005
07/19/2010 09:25午前	ディレクトリ SRA001

DRA: 公開データの量



【実習】DRAsearch で検索してみよう

 DDBJ
DNA Data Bank of Japan

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。

DRAsearch: Arabidopsis を選択

統合データベース × DNA Data Bank × DDBJ Sequence × DRA Search × SRP006839 × SRX065809 × Murasaki Project ×

trace.ddbj.nig.ac.jp/DRASearch/ 他のブックマーク

W DBCLS E P R S Papers Tech travel Diigolet

DRASearch Send Feedback Search Home DRA Home

Accession :

Organism : Ara StudyType :
CenterName : Arabidopsis arenosa Platform :
Keyword : Arabidopsis lyrata
Arabidopsis thaliana
Arachis duranensis
Arachis hypogaea

Show 20 Data Last Update 2012-03-02
WebSite Last Update 2011-06-20

Statistics

Released Entries

Type	Count
Submission	62023
Study	9728
Experiment	123324
Sample	256417
Run	393965

Organism

#	Organism Name	Study
1	Homo sapiens	881
2	unidentified	877
3	Mus musculus	471

Study Type

#	Study Type	Study
1	Whole Genome Sequencing	4879
2	Transcriptome Analysis	1446
3	Metagenomics	1150

Center Name [All List]

#	Center Name	Study
1	JGI	1592
2	GEO	1376
3	JCVI	1310

DRAsearch: RNASeq を選択

The screenshot shows a web browser window for the DRAsearch platform. 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 browser's toolbar includes links to various databases like DNA Data Bank, DDBJ Sequence, and Murasaki Project.

The main search interface includes fields for Accession, Organism (set to *Arabidopsis thaliana*), CenterName, Keyword, and a dropdown for Show records (set to 20) and Sort by (set to Study). Below these are the search buttons "Search" and "Clear".

A prominent feature is a dropdown menu titled "StudyType" which lists various sequencing and genomics methods. The option "RNASeq" is circled in red at the bottom of the list.

- ✓ 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
- RNA-seq
- small RNA
- sorted chromosome sequencing
- subtractive hybridization
- Synthetic Genomics
- Taq ssu rRNA gene hypervariable region taq sequencing

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

DRAsearch: Arabidopsis & RNASeq

Screenshot of the DRAsearch interface showing search results for Arabidopsis thaliana RNASeq studies.

Search parameters:

- Organism: *Arabidopsis thaliana*
- StudyType: RNASeq
- Platform: (dropdown menu)
- Accession: (empty)
- Keyword: (empty)
- Show 20 records
- Sort by Study
- Search
- Clear

Search Results (10 studies):

#	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> 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

The first study (SRP006839) is circled in red.

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

DRASearch: SRP006839

統合データ × DNA Data Bank × DDBJ Sequenc × Result List × SRP006839 × SRP006839 × SRX065809 × Murasaki Project ×

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

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

DRASearch

SRP006839

Study Detail	
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
-------------	-----------------------------------

Navigation

- 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

統合データ × DNA Data × DDBJ Seq. × Result List × SRP00683 × SRX065809 × SRP00683 × SRX065809 × Murasaki ×

trace.ddbj.nig.ac.jp/DRASearch/experiment?acc=SRX065809

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

DRASearch

SRX065809 FASTQ SRALite

Experiment Detail

Title	Longitudinal Sections small RNA sequencing
Design Description	
Organism	Arabidopsis thaliana

Library Description

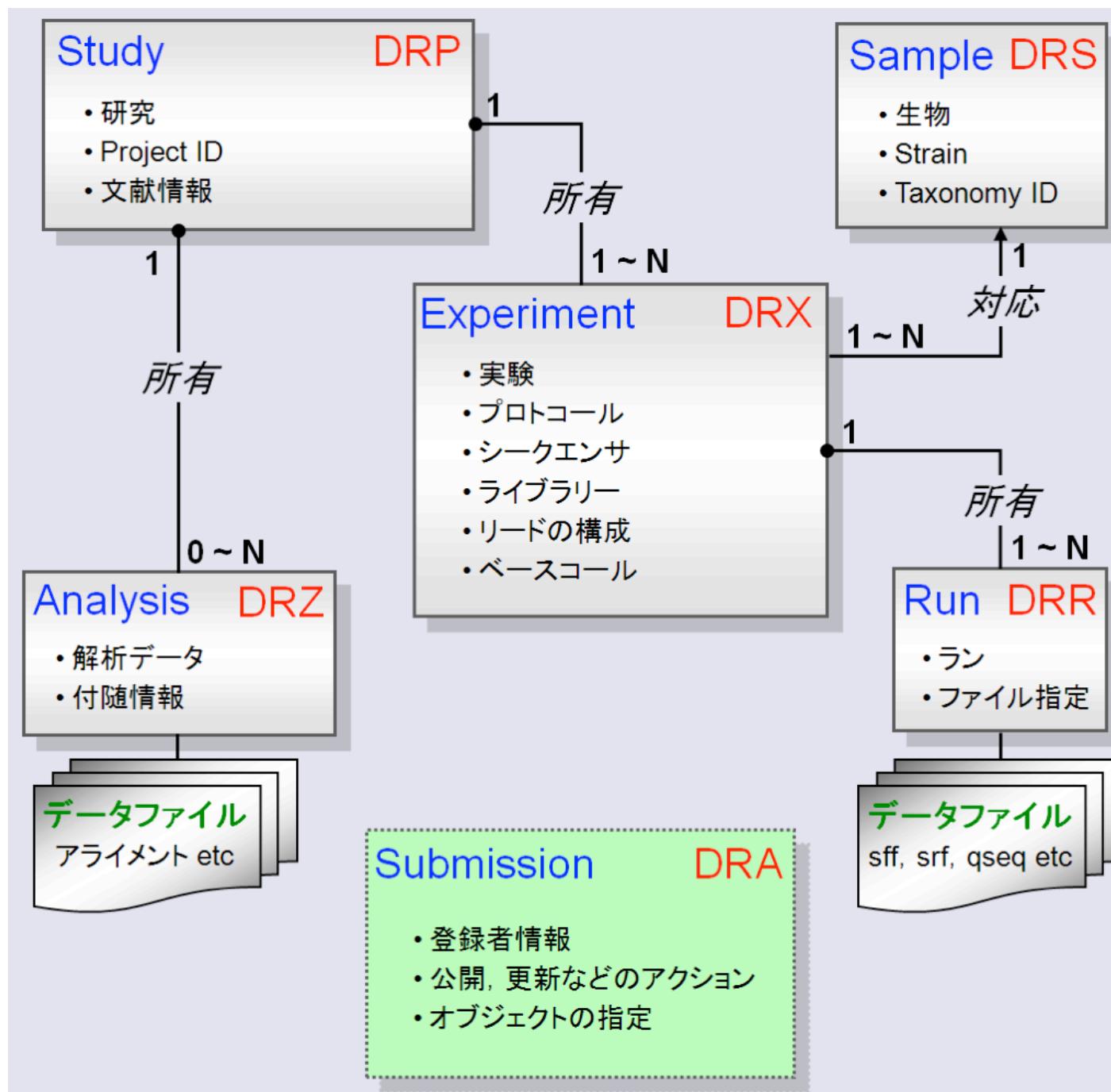
Name	LS1
Strategy	RNA-Seq
Source	TRANSCRIPTOMIC
Selection	size fractionation
Layout	SINGLE

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

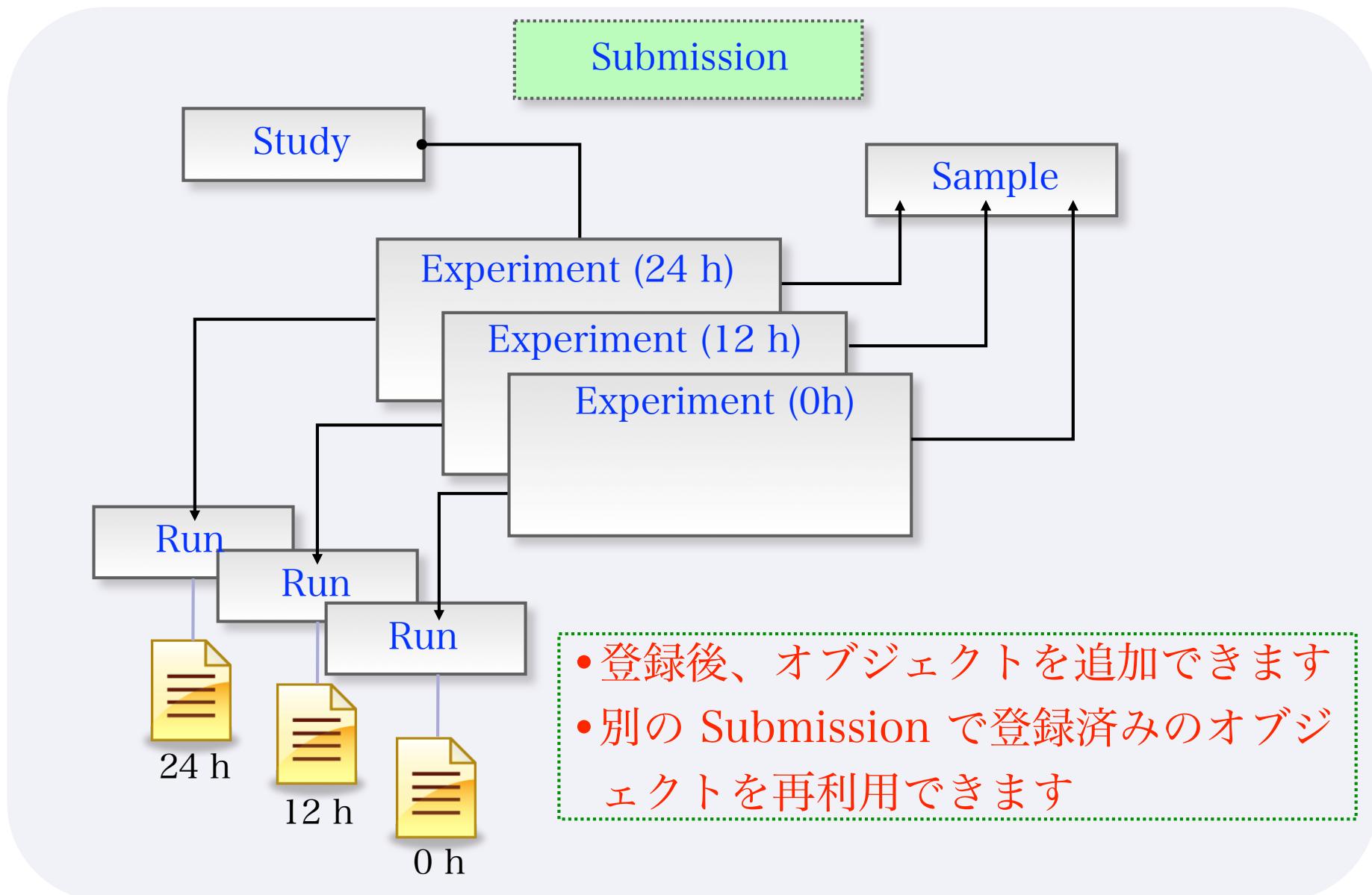
- Submission [SRA037191](#) [FTP](#)
- Study [SRP006839](#)
- Sample [SRS209324](#)
- Run [SRR218085](#) [FASTQ](#) [SRALite](#)
- [SRR218086](#) [FASTQ](#) [SRALite](#)
- [SRR218087](#) [FASTQ](#) [SRALite](#)
- [SRR218088](#) [FASTQ](#) [SRALite](#)

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". The navigation menu includes "Home", "Documentation" (which is circled in red), "Submission", "Search", "Download", "Pipeline", and "About". A language selection "» English" is also present. The main content area starts with a breadcrumb "Home > Metadata" and a numbered list of documentation sections: 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. Below this, a section titled "1. 概要" discusses metadata objects and their relationships. At the bottom, there is a diagram illustrating the relationship between Study, DRP, and Sample DRS.

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

```
graph LR; Study --- DRP[DRP]; DRP --- SampleDRS[Sample DRS];
```

Study

- 研究
- Project ID

DRP

1

Sample DRS

- 生物
- Strain

ネットワークを介して計算機を使い倒す

遺伝研

スーパーコンピュータ



3/21から利用可能になりました！

- <http://www.ddbj.nig.ac.jp/system/supercom/supercom-apl.html>
- [DDBJ スーパーコンピュータ] で検索

The screenshot shows the DDBJ (DNA Data Bank of Japan) website. The header features the DDBJ logo and navigation links for ENGLISH, Twitter, and RSS. Below the header is a search bar and a site search button. The main menu includes links for HOME, 塩基配列の登録, 利用の手引き (highlighted in yellow), 検索・解析, FTP・WebAPI, レポート・統計, and お問い合わせ.

The current page is "HOME > DDBJing目次 : DDBJ利用の手引き > スーパーコンピュータシステムの利用申込". The right sidebar contains a section titled "スーパーコンピュータシステムの利用申込" with text about logging in to the supercomputer system and application procedures. The left sidebar has sections for DDBJ introduction, Q&A, sequence database registration (SAKURA, MSS, etc.), project registration, and search.

スーパーコンピュータシステムの利用申込

国立遺伝学研究所のスーパーコンピュータシステム(以下「スパコン」)にログインして利用する場合は、計算機のアカウントが必要です。情報・システム研究機構国立遺伝学研究所 DDBJ塩基配列データベース等利用規程をご覧の上、利用申請を行って下さい。利用期間は一事業年度です。

利用を継続する場合は、年度末に継続申請の手続きを行って下さい。利用を中止する場合は、国立遺伝学研究所大型計算機利用中止申請を行って下さい。

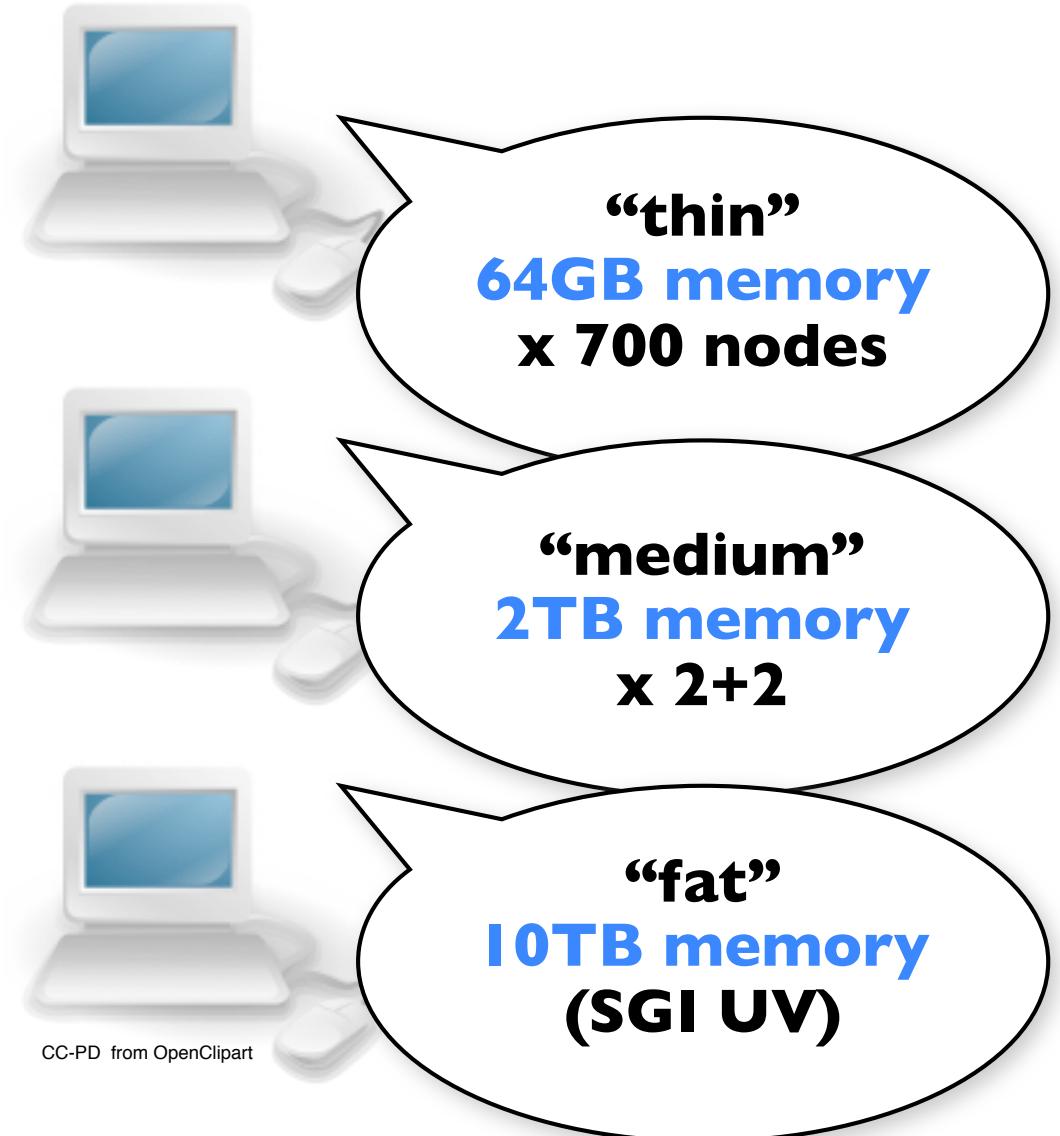
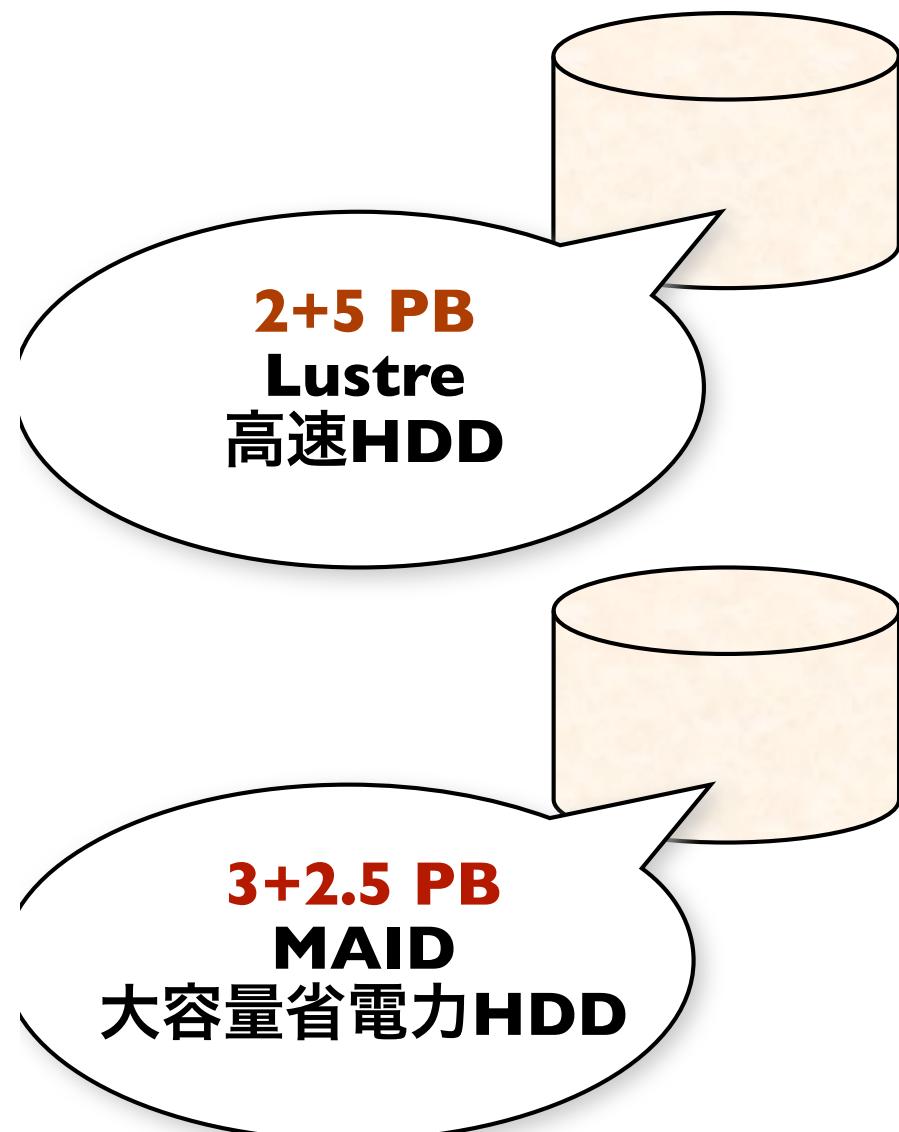
今後もこの方面的サービスの充実を図って参りますので、ご理解とご協力をお願いします。

新規申込

「アカウント新規受付」

スパコンにログインするアカウントを希望の方は、こちらから申込を行ってください。申請受理後、郵送にて登録証を発送いたします。

DDBJ・新スパコン概要 (2012.3)





PROJECT LISTS STATISTICS RESOURCES NEWS

National Institute of Genetics

<http://i.top500.org/site/48477>

URL: <http://www.nig.ac.jp/>
Segment Research
City Mishima
Country Japan

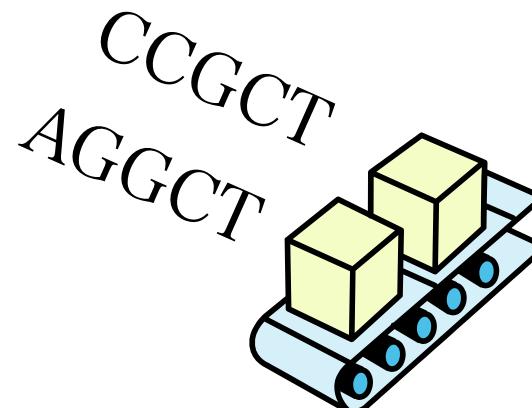
Installed Systems

List	Rank	System	Vendors	Cores	Rmax (GFlop/s)	Rpeak (GFlop/s)
06/2012	280	Cluster Platform SL230s/SL250s, Xeon E5-2670 8C 2.60GHz, Infiniband QDR	Hewlett-Packard	5616	82900	116813
11/2001	357	VPP5000/12	Fujitsu	12	112	115.2
11/2001	389	PRIMEPOWER2000 563 MHz	Fujitsu	128	102	216.2
11/2000	488	VPP500/40	Fujitsu	40	56.9	64
06/2000	262	VPP500/40	Fujitsu	40	56.9	64
11/1999	188	VPP500/40	Fujitsu	40	56.9	64
06/1999	122	VPP500/40	Fujitsu	40	56.9	64
11/1998	93	VPP500/40	Fujitsu	40	56.9	64
06/1998	70	VPP500/40	Fujitsu	40	56.9	64
11/1997	58	VPP500/40	Fujitsu	40	56.9	64
06/1997	47	VPP500/40	Fujitsu	40	56.9	64
11/1996	26	VPP500/40	Fujitsu	40	56.9	64
06/1996	17	VPP500/40	Fujitsu	40	56.9	64
12/1995	14	VPP500/40	Fujitsu	40	56.9	64



DDBJ pipeline

情報処理用語でパイプラインとは、さまざまなプログラムを直列に連結し、工場などの流れ作業のようにプログラムを連続して走らせてデータの処理を高速化・定型化する技術です
新型シーケンサーからのデータを、DDBJのスーパーコンピュータを使って解析できるパイプラインを用意し、莫大なデータを解析するための計算機資源を個別に用意しなくともよく、また、複雑な情報処理技術なしで、各種の解析処理がボタンひとつで簡単に実行できます！



DDBJ Pipeline: クラウド型解析ツール

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

W DBCLS E P R S Papers Tech travel Dilgolet 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-V/GTPS

DDBJ : DNA Data Bank of Japan

DDBJ（日本DNAデータバンク）は欧州と米国との対応機関（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

▶ 一覧へ

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。

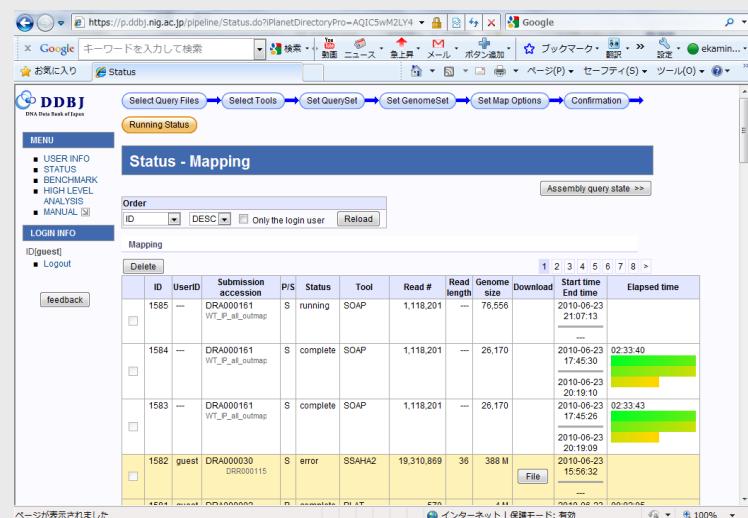
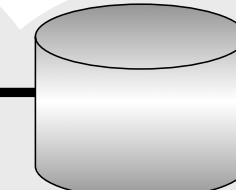
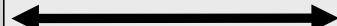
FTP・Web API

■ 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



The screenshot shows a web browser displaying the DDBJ Pipeline status page. The URL is https://p.ddbj.nig.ac.jp/pipeline>Status.do?n=PlanetDirectoryPro-AQIC5wM2LY4. The page title is "Status - Mapping". The left sidebar has a "Status" tab selected. The main content area shows a table of mapping jobs:

ID	User ID	Submission accession	P/S	Status	Tool	Read #	Read length	Genome size	Download	Start time	End time	Elapsed time
1585	---	DRA000161 WT_P_all_outmap	S	running	SOAP	1,118,201	---	76,556		2010-06-23 17:45:30	2010-07-13 02:33:40	2010-06-23 17:45:30
1584	---	DRA000161 WT_P_all_outmap	S	complete	SOAP	1,118,201	---	26,170		2010-06-23 17:45:30	2010-07-10 02:33:43	2010-06-23 17:45:26
1583	---	DRA000161 WT_P_all_outmap	S	complete	SOAP	1,118,201	---	26,170		2010-06-23 17:45:26	2010-07-10 02:33:43	2010-06-23 17:45:26
1582	guest	DRA000030 DRR000115	S	error	SSAHA2	19,310,869	36	388 M	File...	2010-06-23 15:56:32	2010-06-23 02:33:43	2010-06-23 15:56:32

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

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

DRA pipeline: ソフトウェア



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

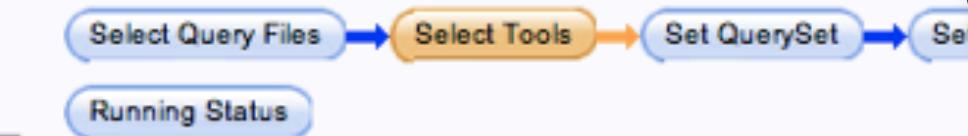
step1-Assembly status

step2-All status

Help

MANUAL

BENCHMARK



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			
<input type="checkbox"/>	BLAT		34	✓			✓	✓	✓			Single-end analysis only
<input type="checkbox"/>	Maq		0.7.1	✓			✓	✓	✓	✓	✓	✓
<input type="checkbox"/>	bwa		0.5.8a	✓			✓	✓	✓			✓
<input type="checkbox"/>	SSAHA2		2.3.0.1	✓			✓			✓	✓	SNP is single-end analysis only
<input type="checkbox"/>	SOAP		2.1.8	✓			✓	✓	✓	✓	✓	✓
<input type="checkbox"/>	Bowtie (SAMtools)		0.12.0 (0.1.7)	✓	✓	✓	✓	✓	✓	✓	✓	✓
<input type="checkbox"/>	TopHat		1.0.11 (BETA)	✓			✓	✓	✓	✓		✓

de novo Assembly

Total limit = 22 Gbp

	Tool	Help	Version	Base space	Color space	Paired-end	MSS(WGS)	Comment
<input type="checkbox"/>	Velvet		0.7.56	✓		✓	✓	
<input type="checkbox"/>	Edena		2.1.1	✓				

feedback

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



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

Running Status

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

feedback

Specifying Database of Reference Genome

RESET BACK NEXT

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

all check

chr01.fasta

chr02.fasta

chr03.fasta

chr04.fasta

chr05.fasta

chr06.fasta

chr07.fasta

chr08.fasta

chr09.fasta

chr10.fasta

chr11.fasta

chr12.fasta

Organisms

Mus musculus

Genome sets

✓ Jul. 2007 (mm9)

Mar. 2006 (mm8)

Aug. 2005 (mm7)

NCBI build 36

NCBI build 37

all check

chr1.fa

chr10.fa

chr11.fa

chr12.fa

Organisms

Arabidopsis thaliana

Genome sets

✓ TAIR8

TAIR9

all check

chr1.fas

chr2.fas

chr3.fas

User original sets

Download or upload reference

DRA pipeline: merits

- DDBJのPC cluster上で運用
- 大量データ転送問題が回避できる
- 最新の *de facto* standard ツールを用意
 - for *de novo* assemble 「アセンブル」
 - for reference genome mapping 「マッピング」
- 処理はおまかせ簡単

SRAsearchで探し、pipelineで処理

- バクテリアのNGSデータをアセンブル
- シロイヌナズナのRNA-seqをTAIRの特定のバージョンにマッピング
- などなど、ご興味のあるNGS公開配列をさがして、トライしてみてください

【実習】基礎処理の流れをつかもう

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



DDBJ Read Annotation Pipeline

English

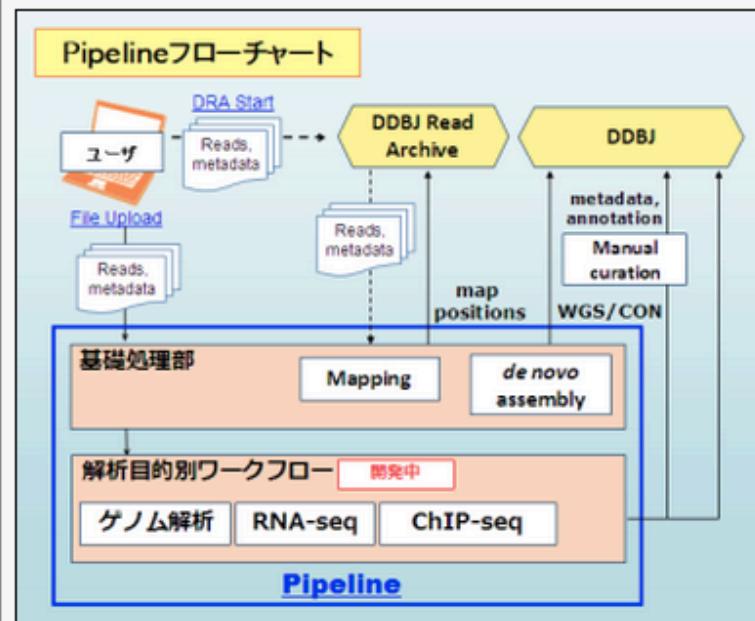
Japanese

DDBJ Read Annotation Pipelineは、次世代シーケンサ配列のクラウド型データ解析プラットフォームです。

LOG IN

新規アカウント作成

ゲストとしてログイン



User ID:

Password:

ゲストとして
ログイン

動作中JOBの確認

PipelineのIDをお持ちでない場合、[ゲストとしてログイン](#)することができます。

マニュアルおよびチュートリアル

- [日本語マニュアル](#)
- [英語マニュアル](#)
- [DBCLS 統合TV チュートリアル1 - 今日からはじめるDDBJ Read Annotation Pipeline](#)
- [DBCLS 統合TV チュートリアル2 - DDBJ Read Annotation Pipelineによるde novo Assembly解析](#)

DRAアカウント登録

DRAアカウントの登録に関しては [こちらをご覧ください](#)。



pipeline_info

pipeline_info !!!Notice!!!! DDBJ pipeline services will
not be available due to system maintenance as

「接続の安全性を確認できません」が出たら

「接続の安全性を確認できません」の画面が出てきたら
セキュリティ例外を承認して下さい。

The screenshot shows a Firefox security warning dialog box. At the top left is a yellow icon of a person holding a shield. The main title is "接続の安全性を確認できません". Below it, a message says: "sso.ddbj.nig.ac.jp に安全に接続するように求められましたが、接続の安全性が確認できませんでした。" A warning message in the center states: "安全に接続する場合は通常、あなたが適切な相手と通信することを確認できるように信頼できます。しかし、このサイトの証明書は信頼性を検証できません。" To the right, a note says: "例外的に信頼する証明書としてこのサイトの証明書を登録しようとしています。本物の銀行、通信販売、その他の公開サイトがこの操作を求めるることはあります。" Below this are sections for "サーバ" (Server) showing the URL "https://sso.ddbj.nig.ac.jp/opensso/DDBJLoginForm?rea" and a "証明書を取得" (Get Certificate) button. The "証明書の状態" (Certificate Status) section notes: "このサイトでは不正な証明書が使用されており、サイトの識別情報を確認できません。" It also says "不明な証明書です" (Unknown certificate) and "既知の認証局によって検証されていないため、このサイトの証明書は信頼されません。" At the bottom, there is a checkbox labeled "次回以降にもこの例外を有効にする" (Keep this exception valid for future sessions) with the checked option, and two buttons: "セキュリティ例外を承認" (Accept security exception) and "キャンセル" (Cancel). A red arrow points from the "例外を追加..." (Add exception...) button in the original browser window to the "セキュリティ例外を承認" button in the dialog box.

処理に使うNGSの配列ファイルの用意

Select Query Files → Select Tools → Options → Confirmation

Running Status

Selecting Query Files

NEXT

FTP upload Private DRA entry Import public DRA Preprocessing HTTP upload

Metadata of the DRA entry.

Select a metadata : DRA000001

TYPE	ACCESSION	ALIAS	FILENAME	DL	VIEW
Submission	DRA000001		DRA000001.submission.xml	DownLoad	View
Sample	DRS000001	Bacillus subtilis subsp. natto BEST195 without plasmid pBEST195L	DRA000001.sample.xml	DownLoad	View
Study	DRP000001	Natto BEST195	DRA000001.study.xml	DownLoad	View
Experiment	DRX000001	NATTO_BEST195_SEP08	DRA000001.experiment.xml	DownLoad	View
Run	DRR000001	2008-09-12.BEST195-Lane7	DRA000001.run.xml	DownLoad	View

STUDY TITLE : Whole genome sequencing of Bacillus subtilis subsp. natto BEST195
STUDY TYPE : Whole Genome Sequencing

Select your registered query files.

Queries with different Instrument models can't be selected together.

single paired all clear

No.	Experiment ACCESSION	Sample ACCESSION	Run ACCESSION	STRAIN	Run_date	Read #	Read length	Instrument model	Layout
1	DRX000001	DRS000001	DRR000001	strain BEST195	2008-09-12	9,977,388	36	ILLUMINA	paired

: from metadata : Counted from query file (Read length is calculated from the first entry.)

DELETE NEXT

処理に使うNGSの配列ファイルの用意

FTP で手元から
アップロード可能

	Filename	Description	Layout	Instrument model	File size
<input type="checkbox"/>	GSM727564_d0Foxh1.bed.gz	Foxh1	single	ILLUMINA	0 byte
<input type="checkbox"/>	unknow1.fastq (more 1 files)	preprocessing	paired	ILLUMINA	48.2 MB
<input type="checkbox"/>	unknow2.fastq	vvvv	single	LS454	27.2 MB
<input type="checkbox"/>	blob (more 1 files)	vivek	paired	ILLUMINA	866.1 MB
<input type="checkbox"/>	blob.1 (more 1 files)	vivek	paired	ILLUMINA	1.5 GB
<input type="checkbox"/>	DRR000985.fastq	123	single	ILLUMINA	3.6 GB
<input type="checkbox"/>	blob (more 1 files)	test	paired	ILLUMINA	866.1 MB

処理に使うNGSの配列ファイルの用意

DDBJ
DNA Data Bank of Japan

ACCOUNT
login ID [guest] Logout

ANALYSIS
Data setup
DRA Start
FTP upload
HTTP upload
DRA Import
Preprocessing Start
step-1
Preprocessing
Mapping /
de novo Assembly
step-2
Workflow
Genome (SNP/Short
Indel)
RNA-seq (Tag count)
ChIP-seq

JOB STATUS
step1.
Preprocessing
step1.
Mapping
step1.
de novo Assembly
step2-All status

HELP
HELP □
TUTORIAL □
Contact Us.
DDBJ Read Annotation
Pipeline.
Development Team.

Select Query Files → Select Tools → Set QuerySet → **Import public DRA** → Confirmation

Running Status

Selecting Query Files

NEXT

公开データを
インポート可能

FTP upload Private DRA entry **Import public DRA** Preprocessing HTTP upload

Import public FASTQ files from DRA database.

Here is do the section of automatic download of public DRA/ERA/SRA entries.
Please input DRA/ERA/SRA accession number. Then the pipeline system import metadata and FASTQ files from DRA database.

Input DRA/ERA/SRA Accession Number
 Add my DRA entry

Accession Number can find here.
[DRA Search](#)

Your request. (Here is display only. can not select.)

To select your downloaded entries. See Private DRA entry tab.
When the status makes "done", your requested entry is added in "Private DRA entry" tabs.
When the status makes "failed" or "preparing", please retry it.

queued : waiting or during download, done : file is ready, failed : please retry it, preparing : file is not yet in DRA unchecked : download is ok, but md5 was not check.

Status	Submission	Request date
✓ done	DRA000001	2013-01-11 18:13:25.174
✗ preparing	SRA060574	2013-01-07 23:49:33.51
✗ preparing	SRA058628	2013-01-07 22:52:08.369
✗ preparing	SRA050143	2012-11-15 19:17:57.271
✗ preparing	SRA046010	2012-10-29 21:50:21.933
✓ done	SRA040340	2012-10-29 15:04:16.249
✓ done	DRA000303	2012-08-27 07:49:30.698
✓ done	DRA000086	2012-08-24 13:51:17.364

NGSデータの前処理: preprocessing

DDBJ
DNA Data Bank of Japan

ACCOUNT
login ID [guest]

ANALYSIS
Data setup
DRA Start
FTP upload
HTTP upload
DRA Import
Preprocessing Start
Preprocessing (Red box)
Mapping
de novo Assembly
step-2
Workflow
Genome (SNP/Short Indel)
RNA-seq (Tag count)
ChIP-seq

JOB STATUS
step1.
Preprocessing
step1.
Mapping
step1.
de novo Assembly
step2-All status

HELP
HELP
TUTORIAL
Contact Us.
DDBJ Read Annotation Pipeline.
Development Team.

Selecting Query Files

FTP upload Private DRA entry

Metadata of the DRA entry.

Select a metadata :

TYPE	ACCESSION	ALIAS	FILENAME	DL	VIEW
Submission	DRA000001	DRA000001	DRA000001.submission.xml	<input type="button" value="DownLoad"/>	<input type="button" value="View"/>
Sample	DRS000001	DRS000001	DRA000001.sample.xml	<input type="button" value="DownLoad"/>	<input type="button" value="View"/>
Study	DRP000001	DRP000001	DRA000001.study.xml	<input type="button" value="DownLoad"/>	<input type="button" value="View"/>
Experiment	DRX000001	DRX000001	DRA000001.experiment.xml	<input type="button" value="DownLoad"/>	<input type="button" value="View"/>
Run	DRR000001	DRR000001	DRA000001.run.xml	<input type="button" value="DownLoad"/>	<input type="button" value="View"/>

STUDY TITLE Whole genome sequencing of *Baillus subtilis* subsp. *natto* BEST195
STUDY TYPE Whole Genome Sequencing

Select your registered query files.

Different instrument models can't be selected together.

No.	Experiment ACCESSION	Sample ACCESSION	Run ACCESSION	STRAIN	Run_date	Read #	Read length	Instrument model	Layout
<input type="checkbox"/>	1 DRX000001	DRS000001	DRR000001	strain BEST195	2008-09-13	9,977,388	36	ILLUMINA	paired

: from metadata : Counted from FASTQ (Sequence length is calculated from the first entry.)

前処理済のデータはここに

先ほどの
ログイン直後の
タブにあり

The screenshot shows the DDBJ analysis interface. At the top, there's a navigation bar with steps: Select Query Files → Select Tools → Set QuerySet → Set GenomeSet. Below this is a 'Running Status' section. The main title is 'Selecting Query Files'. Underneath, there are tabs: FTP upload, Private DRA entry, Import public DRA, Preprocessing (which is highlighted in yellow), and HTTP upload. A large blue arrow points from the left towards the 'Preprocessing' tab. Below these tabs is a section titled 'From Preprocessing output files.' containing a table of files. The table has columns: Filename, Layout, and File size. The 'Preprocessing Start' button in the sidebar is circled in red.

Filename	Layout	File size
3951_DRR000719_1_e.fastq.bz2 (more 1 files)	paired	2.0 GB
3951_DRR000720_1_e.fastq.bz2 (more 1 files)	paired	16.1 GB
4174_DRR000001_1_e.fastq.bz2 (more 1 files)	paired	345.6 MB
4178_DRR000001_1_e.fastq.bz2 (more 1 files)	paired	345.6 MB
4184_4174_DRR000001_1_e_e.fastq.bz2 (more 1 files)	paired	345.6 MB
4184_4178_DRR000001_1_e_e.fastq.bz2 (more 1 files)	paired	345.6 MB
4185_4174_DRR000001_1_e_e.fastq.bz2 (more 1 files)	paired	345.6 MB
4185_4178_DRR000001_1_e_e.fastq.bz2 (more 1 files)	paired	345.6 MB
4519_DRR000001_1_e.fastq.bz2 (more 1 files)	paired	345.6 MB
4523_DRR000001_1.unmapped.fastq_4337.bz2 (more 1 files)	paired	821.6 MB
4652_DRR000001_1.unmapped.fastq_4466.bz2 (more 1 files)	paired	28 byte

DELETE NEXT

ACCOUNT
login ID [guest] Logout

ANALYSIS
Data setup
DRA Start
FTP upload
HTTP upload
DRA Import
Preprocessing Start

step-1
Preprocessing
Mapping /
de novo Assembly

step-2
Workflow
Genome (SNP/Short
Indel)
RNA-seq (Tag count)
ChIP-seq

JOB STATUS
step1.
Preprocessing
step1.
Mapping
step1.
de novo Assembly
step2-All status

HELP
HELP
TUTORIAL
Contact Us.
DDBJ Read Annotation
Pipeline.

今回はupload済のエントリから

DDBJ
DNA Data Bank of Japan

ACCOUNT
login ID [guest] Logout

ANALYSIS
Data setup
DRA Start
FTP upload
HTTP upload
DRA Import
Preprocessing Start
step-1
Preprocessing
Mapping /
de novo Assembly
step-2
Workflow
Genome (SNP/Short
Indel)
RNA-seq (Tag count)
ChIP-seq

JOB STATUS
step1.
Preprocessing
step1.
Mapping
step1.
de novo Assembly
step2-All status

HELP
HELP ☰
TUTORIAL ☰
Contact Us.
DDBJ Read Annotation
Pipeline.
Development Team.

Select Query Files → Select Tools → Set QuerySet → Set GenomeSet → Set Map Options → Confirmation →

Running Status

NEXT

Selecting Query Files

FTP upload Private DRA entry Import public DRA Preprocessing HTTP upload

Metadata of the DRA entry.

Select a metadata file

TYPE	ACCESSION	ALIAS	FILENAME	DL
Submission	DRA000001	DRA000001	DRA000001.submission.xml	
Sample	DRS000001	DRS000001	DRA000001.sample.xml	
Study	DRP000001	DRP000001	DRA000001.study.xml	
Experiment	DRX000001	DRX000001	DRA000001.experiment.xml	
Run	DRR000001	DRR000001	DRA000001.run.xml	

STUDY TITLE Whole genome sequencing of *Baillus subtilis* subsp. *natto* BEST195
STUDY TYPE Whole Genome Sequencing

Select your registered query files.

Queries with different Instrument models can't be selected together.

single paired all clear

No.	Experiment ACCESSION	Sample ACCESSION	Run ACCESSION	STRAIN	Run_date	Read #	Read length	Instrument model	Layout
1	DRX000001	DRS000001	DRR000001	strain BEST195	2008-09-13	9,977,388	36	ILLUMINA	paired

: from metadata : Counted from query file (Read length is calculated from the first entry.)

DELETE NEXT

納豆菌の
公開データが
インポート済

ACCOUNT
login ID [guest]
Logout

ANALYSIS
Data setup
DRA Start
FTP upload
HTTP upload
DRA Import
Preprocessing Start
step-1
Preprocessing
Mapping / <i>de novo</i> Assembly
step-2
Workflow
Genome (SNP/Short Indel)
RNA-seq (Tag count)
ChIP-seq

JOB STATUS
step1. Preprocessing
step1. Mapping
step1. <i>de novo</i> Assembly
step2-All status

HELP
HELP
TUTORIAL
Contact Us. DDBJ Read Annotation Pipeline. Development Team.

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		0.5.9	✓			✓			✓				✓	
SOAP		2.21	✓			✓			✓	✓	✓			
Bowtie		0.12.7	✓	✓	✓				✓	✓				
TopHat		1.0.11	✓			✓			✓					
Bowtie2		2.0.0	✓	✓	✓				✓	✓				For reads longer than about 100 bp

de novo Assembly

Total limit = 22 Gbp

Tool	Help	Version	Base space	Color space	Paired-end	MSS(WGS)	Comment
SOAPdenovo		1.05			✓		
ABySS		1.3.2			✓		Maximum K-mer value is 64.
<input checked="" type="checkbox"/> Velvet		1.2.03			✓	✓	We severely recommend when performing Velvet, total length of those reads is up to 22G bp. Maximum K-mer value is 64.
Trinity		r2012-06-08			✓		RNA-Seq De novo Assembly

Mapping Contigs by *de novo* Assemble to Reference Sequences.

The contigs will be aligned to reference genome.

Tool	Comment
BLAT	Single-end analysis only

velvet で
アセンブル
しましょう

[BACK](#) [NEXT](#)

配列とペア形式のセットを選択

DDBJ DNA Data Bank of Japan

ACCOUNT
login ID [guest] Logout

ANALYSIS
Data setup
DRA Start
FTP upload
HTTP upload
DRA Import
Preprocessing Start
step-1
Preprocessing
Mapping /
de novo Assembly
step-2
Workflow
Genome (SNP/Short
Indel)
RNA-seq (Tag count)
ChIP-seq

JOB STATUS
step1.
Preprocessing
step1.
Mapping
step1.
de novo Assembly
step2-All status

HELP
HELP
TUTORIAL
Contact Us.
DDBJ Read Annotation
Pipeline.
Development Team.

Select Query Files → Select Tools → Set QuerySet → Set Ass. Options → Confirmation → Running Status

Generating Query Sets from Query Read Files

RESET BACK NEXT

Paired-end analysis
Layout of paired sequence. 5'-3' 3'-5'
5' 3' 3' 5'
Linker(1) Target Linker(2) Linker(3) Target Linker(4)

	Run ACCESSION	Read length	Quality Score
<input checked="" type="checkbox"/>	D0R000001 -><-	36 bp	Read1 Read2

Set as Mate-Pair Set as Pair-End

QUERY SET

RESET BACK NEXT

配列のセットの形式を選んで次へ

DDBJ DNA Data Bank of Japan

Select Query Files → Select Tools → Set QuerySet → Set Ass. Options → Confirmation → Running Status

ACCOUNT
login ID [guest] Logout

ANALYSIS
Data setup
DRA Start
FTP upload
HTTP upload
DRA Import
Preprocessing Start
step-1
Preprocessing
Mapping /
de novo Assembly
step-2
Workflow
Genome (SNP/Short
Indel)
RNA-seq (Tag count)
ChIP-seq

JOB STATUS
step1.
Preprocessing
step1.
Mapping
step1.
de novo Assembly
step2-All status

HELP
HELP ▾
TUTORIAL ▾
Contact Us.
DDBJ Read Annotation
Pipeline.
Development Team.

Generating Query Sets from Query Read Files

Paired-end analysis
Layout of paired sequence. 5'-3' 3'-5'
5' 3' 3' 5'
Linker(1) Target Linker(2) Linker(3) Target Linker(4)

Run ACCESSION Read length Quality Score

Set as Mate-Pair Set as Pair-End

QUERY SET
Query set1

PairedOrientation	RunAccession	RunAlias	RowLength	QualityScore1	QualityScore2
paired	DRR000001	DRR000001	36		

RESET BACK NEXT

オプションのパラメータを選べます

DDBJ DNA Data Bank of Japan

ACCOUNT
login ID [guest] Logout

ANALYSIS
Data setup
DRA Start
FTP upload
HTTP upload
DRA Import
Preprocessing Start
step-1
Preprocessing
Mapping / de novo Assembly
step-2
Workflow
Genome (SNP/Short Indel)
RNA-seq (Tag count)
ChIP-seq

JOB STATUS
step1. Preprocessing
step1. Mapping
step1. de novo Assembly
step2-All status

HELP
HELP □
TUTORIAL □
Contact Us.
DDBJ Read Annotation Pipeline.
Development Team.

Select Query Files → Select Tools → Set QuerySet → Set Ass. Options → Confirmation → Running Status

Setting for De Novo Assembly

velvet

Set optional parameters of the paired-end analysis

Step1) Convert sequences

Shuffle the sequence.
perl shuffleSequences_fastq.pl query_1.fastq query_2.fastq shuffle_query_pe.fastq

Running velvet.
Velveth output_directory/ -shortPaired shuffle_query_pe.fastq

Step2) Assembly

Velvetg output_directory/

Step3) Set parameters of the CONFIG mapping tool

Step4) Create assembled sequences in FASTA file from pileupped reads to [submit WGS division of DDBJ](#).

Set filtered length for contigs
 perl lengthfilter.pl pileupFile out_WGS.txt

BACK **NEXT**

特になければ
そのまま次へ

終了したらメールが来ます

DDBJ
DNA Data Bank of Japan

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

ANALYSIS
Data setup
DRA Start
FTP upload
HTTP upload
DRA Import
Preprocessing Start
step-1
Preprocessing
Mapping /
de novo Assembly
step-2
Workflow
Genome (SNP/Short
Indel)
RNA-seq (Tag count)
ChIP-seq

JOB STATUS
step1.
Preprocessing
step1.
Mapping
step1.
de novo Assembly
step2-All status

HELP
[HELP](#) [TUTORIAL](#)
[Contact Us](#) DDBJ Read Annotation Pipeline
Development Team

Select Query Files → Select Tools → Set QuerySet → Set Analysis → Run Confirmation

Run Confirmation

Destination of mail
When the request is completed, the system sends an email to this address:
yn@nig.ac.jp

連絡先いれたら
実行可能

でも今は
押さないで！

Assembly [velvet]

Query sets

Query set1	PairedOrientation	RunAccession	RunAlias	RowLength	QualityScore1	QualityScore2
paired	DRR000001	DRR000001	36			

Assembly commands

velvet

Set optional parameters of the paired-end analysis

Step1) Convert sequences

Shuffle the sequence.
perl shuffleSequences_fastq.pl query_1.fastq query_2.fastq shuffle_query_pe.fastq

Running velvet.

Velveth output_directory/ -shortPaired shuffle_query_pe.fastq

Step2) Assembly

Velvetg output_directory/

Step3) Set parameters of the CONFIG mapping tool

Step4) Create assembled sequences in FASTA file from pileupped reads to [submit WGS division of DDBJ](#).

Set filtered length for contigs -readLengthFilterFile 100 out_WGS.fasta

「RUN を押した」と思ってください

処理状況は
こちらから

The screenshot shows the DDBJ de novo Assembly status page. On the left, a sidebar displays account information (login ID [guest], Logout) and analysis options (Data setup, DRA Start, FTP upload, HTTP upload, DRA Import, Preprocessing Start). Below this is a 'JOB STATUS' section with a red circle around 'step1. de novo Assembly'. The main content area is titled 'Status - de novo Assembly' and includes a navigation bar with tabs: Mapping Job (blue), de novo Assembly Job (orange, selected), and Preprocessing Job (blue). A progress bar at the top indicates the process flow: Select Query Files → Select Tools → Set QuerySet → Set Ass. Options → Confirmation → Running Status. The main table lists assembly jobs with columns for ID, UserID, Submission accession, P/S, Status, Tool, Read #, Read length, Assembly detail, Mapping detail, Start time, End time, and Elapsed time. The 'Assembly detail' column contains a 'View' button, which is also highlighted with a red circle.

ID	UserID	Submission accession	P/S	Status	Tool	Read #	Read length	Assembly detail	Mapping detail	Start time	End time	Elapsed time
4914	guest	DRA000001 DRR000001	P	complete	SOAPdenovo	9,977,388	36	View		2013-01-11 22:06:40	2013-01-11 22:13:40	00:06:59
4912	guest	---	S	complete	SOAPdenovo	1	--	View		2013-01-11 18:01:34	2013-01-11 18:02:16	00:00:42
4911	guest	---	S	error	SOAPdenovo		--	View		—	—	
4909	---	HPS1	P	complete	ABYSS	5,754,246	--			2013-01-11 12:14:34	2013-01-11 13:52:21	01:37:47
4908	guest	DRA000001 2008-09-12.BES	P	complete	Velvet	9,977,388	36	View		2013-01-11 12:13:53	2013-01-11 16:10:01	03:56:08
4907	---	HPS1	P	complete	ABYSS	5,754,246	--			2013-01-11 12:03:54	2013-01-11 13:24:40	01:20:45
4900	---	Ion_mt20mergec	S	complete	ABYSS	148,666	--			2013-01-10 15:32:13		00:16:28

ACCOUNT

login ID [guest]

Logout

ANALYSIS

Data setup

DRA Start

FTP upload

HTTP upload

DRA Import

Preprocessing Start

step-1

Preprocessing

Mapping /
de novo Assembly

step-2

Workflow

Genome (SNP/Short

Indel)

RNA-seq (Tag count)

ChIP-seq

JOB STATUS

step1.

Preprocessing

step1.

Mapping

step1.

de novo Assembly

step2-All status

HELP

HELP

TUTORIAL

Contact Us.

DDBJ Read Annotation
Pipeline.

Development Team.

Detail view

BACK

Job info**ID**

4914

Tool (Version)

SOAPdenovo (1.05)

RunAccession or Filename	Download	Read length	Alias
DRR000001	DRR000001.fastq.gz DRR000001_1.fastq.gz DRR000001_2.fastq.bz2	36 bp	DRR000001

Download modified queries

- [DRR000001_1.fastq.gz \(Original size 1.7 GB\)](#)
- [DRR000001_2.fastq.gz \(Original size 1.7 GB\)](#)

Download wgs file

- [out_WGS.fasta.gz \(Original size 3.9 MB\)](#)

Assembly statistics

Contig # : 5,300
 Total contig size : 4,138,179
 Maximum contig size : 49,938
 Minimum contig size : 24
 N50 contig size : 13,255

アセンブル結果の
基本情報

Time

Wait time	Start time	End time
0: 1:22	2013-01-11 22:06:40	2013-01-11 22:13:40

Command	Start time	End time	Log1	Log2	Result	MD5
SOAPdenovo127mer all -s soapdenovo.conf -o output	2013-01-11 22:06:40	2013-01-11 22:12:28	View	View	Download(174.7 MB)	MD5

結果ファイル

BACK

次はMappingの例: これをロード

DRA Search Send Feedback [Search Home](#) [DRA Home](#)

Accession :

Organism :

Arabidopsis thaliana

StudyType :

RNASeq

CenterName :

Platform :

ILLUMINA

シロイヌナズナ
の RNA-Seq

STUDY_ID	STUDY_TITLE	STUDY_TYPE	ORGANISM	BASES	SUBMITTED	CENTER_NAME
SRP000191	High-resolution profiling of small RNAs in the Arabidopsis thaliana root	RNASeq	Arabidopsis thaliana	9.2G	2011-05-24	Duke University
2 SRP007763	Genome-wide detection of context-sensitive alternative splicing in Arabidopsis roots	RNASeq	Arabidopsis thaliana	1.3G	2011-08-11	Institute of Plant and Microbial Biology, Academia
3 SRP007845	Genome-wide detection of context-sensitive alternative splicing in Arabidopsis roots	RNASeq	Arabidopsis thaliana	2.3G	2011-08-18	Institute of Plant and Microbial Biology, Academia
4 SRP008262	GSE32216: SKIP Is a Splicing Factor Linking Alternative Splicing and Circadian Clock in Arabidopsis	RNASeq	Arabidopsis thaliana	3.7G	2011-09-19	GEO
5 SRP008348	GSE32318: 2-week-old Arabidopsis seedlings (Columbia ecotype)	RNASeq	Arabidopsis thaliana	3.4G	2011-09-23	GEO
6 SRP008486	GSE32284: IBM1, a JmjC domain histone demethylase, is involved in the regulation of RNA-directed DNA methylation through epigenetic control of RDR2 and DCL3 expression in Arabidopsis.	RNASeq	Arabidopsis thaliana	3.6G	2011-09-28	GEO
7 SRP008822	Unexpected diversity of chloroplast non-coding RNAs as revealed by deep sequencing of the Arabidopsis transcriptome	RNASeq	Arabidopsis thaliana	9.8G	2011-10-12	Salk-E
8 SRP009136	Alternative splicing landscape in Arabidopsis	RNASeq	Arabidopsis thaliana	17.6G	2011-10-28	MUW
9 SRP009340	GSE33713: Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos	RNASeq	Arabidopsis thaliana	7.7G	2011-11-15	GEO
10 SRP009369	Mapping Gene Activity of Arabidopsis Root Hairs	RNASeq	Arabidopsis thaliana	22.7G	2011-11-17	Institute of Plant and Microbial Biology, Academia
11 SRP009850	GSE34476: Analysis of the Arabidopsis shoot meristem transcriptome during floral transition identifies distinct cell cycle and meristematic programs	RNASeq	Arabidopsis thaliana	3.8G	2011-12-15	GEO

【実習】もう一度最初から

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



DDBJ Read Annotation Pipeline

English

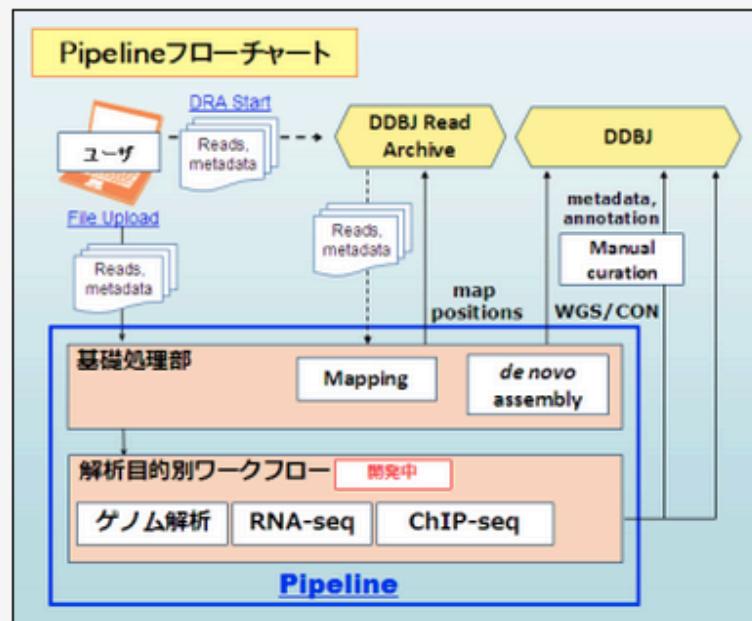
Japanese

DDBJ Read Annotation Pipelineは、次世代シーケンサ配列のクラウド型データ解析プラットフォームです。

LOG IN

新規アカウント作成

ゲストとしてログイン



User ID:

Password:

ゲストとして
ログイン

動作中JOBの確認

PipelineのIDをお持ちでない場合、[ゲストとしてログインすることができます](#)。

マニュアルおよびチュートリアル

- [日本語マニュアル](#)
- [英語マニュアル](#)
- [DBCLS 統合TV チュートリアル1 - 今日からはじめるDDBJ Read Annotation Pipeline](#)
- [DBCLS 統合TV チュートリアル2 - DDBJ Read Annotation Pipelineによるde novo Assembly解析](#)

DRAアカウント登録

DRAアカウントの登録に関しては [こちらをご覧ください](#)。



pipeline_info

pipeline_info !!!Notice!!!! DDBJ pipeline services will
not be available due to system maintenance as

さっき検索したシロイヌナズナのRNA-Seq

DDBJ
DNA Data Bank of Japan

ACCOUNT
login ID [guest] Logout

ANALYSIS
Data setup
DRA Start
FTP upload
HTTP upload
DRA Import
Preprocessing Start
step-1
Preprocessing
Mapping /
de novo Assembly
step-2

Workflow
Genome (SNP/Short
Indel)
RNA-seq (Tag
ChIP-seq)

JOB STATUS
step1.
Preprocessing
step1.
Mapping
step1.
de novo Assembly
step2-All status

HELP
HELP
TUTORIAL
Contact Us.
DDBJ Read Annotation
Pipeline.
Development Team.

Select Query Files → Select Tools → Set QuerySet → Set GenomeSet → Set Map Options → Confirmation → Running Status

Selecting Query Files

FTP upload Private DRA entry Import public DRA Preprocessing HTTP upload

Import public FASTQ files from DRA database.

Here is do the section of automatic download of public DRA/ERA/SRA entries.
Please input DRA/ERA/SRA accession number. Then the pipeline system import metadata and FASTQ files from DRA database.

Input DRA/ERA/SRA Accession Number
SRA044892 Add my DRA entry

Accession Number can find here.
[DRA Search](#)

でも今は押さないで！

an not select.)
e DRA entry tab.
entry is added in "Private DRA entry" tabs.
please retry it.

Step 1: Preprocessing
Step 2: Mapping
Step 3: de novo Assembly
Step 4: All status

Status	Submission	Request date
queued	SRA044892	2013-01-11 23:19:32.978
done	DRA000001	2013-01-11 18:13:25.174
preparing	SRA060574	2013-01-07 23:49:33.51
preparing	SRA058628	2013-01-07 22:52:08.369
preparing	SRA050143	2012-11-15 19:17:57.271
preparing	SRA046010	2012-10-29 21:50:21.933
done	SRA040340	2012-10-29 15:04:16.249
done	DRA000303	2012-08-27 07:49:30.698
done	DRA000086	2012-08-24 13:51:17.364
done	DRA000582	2012-08-20 14:24:46.08

あらかじめ、ロードしておきました

DDBJ DNA Data Bank of Japan

ACCOUNT
login ID [guest] Logout

ANALYSIS
Data setup
DRA Start
FTP upload
HTTP upload
DRA Import
Preprocessing Start
step-1
Preprocessing
Mapping /
de novo Assembly
step-2
Workflow
Genome (SNP/Short
Indel)
RNA-seq (Tag count)
ChIP-seq

JOB STATUS
step1.
Preprocessing
step1.
Mapping
step1.
de novo Assembly
step2-All status

HELP
HELP □
TUTORIAL □
Contact Us.
DDBJ Read Annotation
Pipeline.
Development Team.

Select Query Files → Select Tools → Set QuerySet → Set GenomeSet → Set Map Options → Confirmation →

Running Status

Selecting Query Files

NEXT

FTP upload Private DRA entry Import public DRA Preprocessing HTTP upload

Metadata of the DRA entry.

Select a metadata SRA044892

TYPE	ACCESSION	ALIAS	FILENAME	DL	VIEW
Submission	SRA044892	AraEsFeP	SRA044892.submission.xml	DownLoad	View
Sample	SRS256250 SRS256251 SRS256252	Control Iron Phosphate	SRA044892.sample.xml	DownLoad	View
Study	SRP007763	Alternative Splicing	SRA044892.study.xml	DownLoad	View
Experiment	SRX092046	control	SRA044892.experiment.xml	DownLoad	View
Run	SRR331219 SRR331224	control iron	SRA044892.run.xml	DownLoad	View

STUDY TITLE: Genome-wide detection of context-sensitive alternative splicing in *Arabidopsis* roots
STUDY TYPE: RNASeq

Select your registered query files.

Queries with different Instrument models can't be selected together.

single paired all clear

No.	Experiment ACCESSION	Sample ACCESSION	Run ACCESSION	STRAIN	Run_date	Read #	Read length	Instrument model	Layout
<input type="checkbox"/>	1 SRX092046	SRS256250	SRR331219					ILLUMINA	single
<input type="checkbox"/>	2 SRX092046	SRS256250	SRR331224					ILLUMINA	single

: from metadata : Counted from query file (Read length is calculated from the first entry.)

ANALYSIS

Data setup

DRA Start

FTP upload

HTTP upload

DRA Import

Preprocessing Start

step-1

Preprocessing

Mapping /
de novo Assembly

step-2

Workflow

Genome (SNP/Short
Indel)
RNA-seq (Tag count)
ChIP-seq

JOB STATUS

step1.
Preprocessingstep1.
Mappingstep1.
de novo Assembly

step2-All status

HELP

HELP

TUTORIAL

 Contact Us.
DDBJ Read Annotation
Pipeline.
Development Team.

Selecting Query Files

NEXT

FTP upload

Private DRA entry

Import public DRA

Preprocessing

HTTP upload

Metadata of the DRA entry.

Select a metadata : SRA044892

TYPE	ACCESSION	ALIAS	FILENAME	DL	VIEW
Submission	SRA044892	AraEsFeP	SRA044892.submission.xml	DownLoad	View
Sample	SRS256250 SRS256251 SRS256252	Control Iron Phosphate	SRA044892.sample.xml	DownLoad	View
Study	SRP007763	Alternative Splicing	SRA044892.study.xml	DownLoad	View
Experiment	SRX092046	control	SRA044892.experiment.xml	DownLoad	View
Run	SRR331219 SRR331224	control iron	SRA044892.run.xml	DownLoad	View

STUDY TITLE Genome-wide detection of context-sensitive alternative splicing in Arabidopsis roots

STUDY TYPE RNASeq

Select your registered query files.

Queries with different Instrument models can't be selected together.

single paired all clear

No.	Experiment ACCESSION	Sample ACCESSION	Run ACCESSION	STRAIN	Run_date	Read #	Read length	Instrument model	Layout
<input checked="" type="checkbox"/> 1	SRX092046	SRS256250	SRR331219					ILLUMINA	single
<input type="checkbox"/> 2	SRX092046	SRS256250	SRR331224					ILLUMINA	single

: from metadata

: Counted from query file (Read length is calculated from the first entry.)

DELETE

NEXT

Bowtie2 を選んで NEXT

DDBJ DNA Data Bank of Japan

ACCOUNT
login ID [guest] Logout

ANALYSIS
Data setup
DRA Start
FTP upload
HTTP upload
DRA Import
Preprocessing Start
step-1
Preprocessing
Mapping / de novo Assembly
step-2
Workflow
Genome (SNP/Short Indel)
RNA-seq (Tag count)
ChIP-seq

JOB STATUS
step1. Preprocessing
step1. Mapping
step1. de novo Assembly
step2-All status

HELP
HELP TUTORIAL Contact Us. DDBJ Read Annotation Pipeline. Development Team.

Select Query Files → Select Tools → Set QuerySet → Set GenomeSet → Set Map Options → Confirmation

Running Status

Selecting Tools for Basic Analysis of DDBJ ANNOTATION PIPELINE

BACK **NEXT**

① Reference Genome Mapping

	Tool	Help	Version	Base space	Color space	Paired end	Depth	Coverage	Error rate	SNP	Indel	.gff	.bed	SAM	Comment
<input type="checkbox"/>	BLAT		34	✓					✓						Single-end analysis only
<input type="checkbox"/>	Mag		0.7.1	✓			✓		✓	✓	✓	✓	✓	✓	
<input type="checkbox"/>	bwa		0.5.9	✓			✓		✓					✓	
<input type="checkbox"/>	SOAP		2.21	✓			✓		✓	✓	✓	✓		✓	
<input type="checkbox"/>	Bowtie		0.12.7	✓	✓	✓			✓	✓	✓			✓	
<input type="checkbox"/>	TopHat		1.0.11	✓		✓			✓					✓	

Bowtie2 2.0.0 ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ For reads longer than about 50 bp, Bowtie2 is generally faster, more sensitive, and uses less memory than Bowtie1.

② de novo Assembly
Total limit = 22 Gbp

Tool	Help	Version	Base space	Color space	Paired-end	MSS(WGS)	Comment	
<input type="checkbox"/>	SOAPdenovo		1.05			✓		
<input type="checkbox"/>	ABYSS		1.3.2			✓		Maximum K-mer value is 64.
<input type="checkbox"/>	Velvet		1.2.03			✓	✓	We severely recommend when performing Velvet, total length of those reads is up to 22G bp. Maximum K-mer

配列を選んで confirm, NEXT

DDBJ DNA Data Bank of Japan

ACCOUNT
login ID [guest] Logout

ANALYSIS
Data setup
DRA Start
FTP upload
HTTP upload
DRA Import
Preprocessing Start
step-1
Preprocessing
Mapping /
de novo Assembly
step-2

Workflow
Genome (SNP/Short
Indel)
RNA-seq (Tag count)
ChIP-seq

JOB STATUS
step1.
Preprocessing
step1.
Mapping
step1.
de novo Assembly
step2-All status

HELP
HELP
TUTORIAL
Contact Us.
DDBJ Read Annotation
Pipeline.
Development Team.

Select Query Files → Select Tools → Set QuerySet → Set GenomeSet → Set Map Options → Confirmation

Running Status

Generating Query Sets from Query Read Files

RESET BACK NEXT

Single analysis
Layout of single sequence.
5' Linker(1) Target Linker(2) 3'

Run	ACCESSION	Read length	Quality Score
<input checked="" type="checkbox"/>	SRR331219 ->	bp	

confirm

QUERY SET

RESET BACK NEXT

TAIR10（最新）を選んでNEXT

DDBJ
DNA Data Bank of Japan

ACCOUNT
login ID [guest] Logout

ANALYSIS
Data setup
DRA Start
FTP upload
HTTP upload
DRA Import
Preprocessing Start
step-1
Preprocessing
Mapping /
de novo Assembly
step-2
Workflow
Genome (SNP/Short
Indel)
RNA-seq (Tag count)
ChIP-seq

JOB STATUS
step1.
Preprocessing
step1.
Mapping
step1.
de novo Assembly
step2-All status

HELP
HELP
TUTORIAL
Contact Us.
DDBJ Read Annotation
Pipeline.
Development Team.

Select Query Files → Select Tools → Set QuerySet → Set GenomeSet → Set Map Options → Confirmation →

Running Status

Specifying Database of Reference Genome

RESET BACK NEXT

Major genome sets

Organisms: Arabidopsis thaliana

Genome sets:

- TAIR8
- TAIR9
- TAIR10

all check

- chr01.fa
- chr02.fa
- chr03.fa
- chr04.fa
- chr05.fa
- chrC.fa
- chrM.fa

User original sets

Download or upload reference

RESET BACK NEXT

option 変更なければそのままNEXT

DDBJ DNA Data Bank of Japan

ACCOUNT
login ID [guest] Logout

ANALYSIS
Data setup
DRA Start
FTP upload
HTTP upload
DRA Import
Preprocessing Start
step-1
Preprocessing
Mapping /
de novo Assembly
step-2
Workflow
Genome (SNP/Short
Indel)
RNA-seq (Tag count)
ChIP-seq

JOB STATUS
step1.
Preprocessing
step1.
Mapping
step1.
de novo Assembly
step2-All status

HELP
HELP
TUTORIAL
Contact Us.
DDBJ Read Annotation
Pipeline.
Development Team.

Select Query Files → Select Tools → Set QuerySet → Set GenomeSet → Set Map Options → Confirmation →

Running Status

Setting for Reference Genome Mapping

BACK **NEXT**

bowtie2
Set optional parameters of the single-end analysis

Step1) Convert reference sequence
bowtie2-build -f refgenome.fasta bt2-idx

Step2) Map
bowtie2 -q -p 4 -x bt2-idx -U query1.fastq(fasta) -S out.sam --u out.unmapped

Step3) Convert the read alignment to .BAM format
samtools view -bS -o out.bam out.sam

Step4) Detect DNA polymorphism
Please choose one of the following.

samtools pileup -c -c -f refgenome.fasta out.bam | bcftools view

samtools mpileup -u -C50 -BQ0 -d10000000 -f refgenome.fasta out.bam | bcftools view -bvcg -> out.var.raw.bcf

bcftools view out.var.raw.bcf | vcftools pl varFilter -D10000 > out.var.flt.vcf

Step5) Analysis for Depth, Coverage
samtools sort -o out.bam out_sorted.bam
samtools pileup -c -f reference.fa out_sorted.bam > out.pileup
perl pileup_for_CoverageDepth.pl out.pileup reference.fa
* This command does not appear in the list.

Step6) Create assembled sequences in FASTA file from pileupped reads to submit WGS division of DDBJ.

終了したらメールが来ます

DDBJ
DNA Data Bank of Japan

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

ANALYSIS
Data setup
[DRA Start](#)
[FTP upload](#)
[HTTP upload](#)
[DRA Import](#)
[Preprocessing Start](#)
step-1
[Preprocessing](#)
[Mapping /
de novo Assembly](#)
step-2
Workflow
[Genome \(SNP/Short
Indel\)](#)
[RNA-seq \(Tag count\)](#)
[ChIP-seq](#)

JOB STATUS
step1.
[Preprocessing](#)
step1.
[Mapping](#)
step1.
[de novo Assembly](#)
step2-All status

HELP
[HELP](#)
[TUTORIAL](#)
[Contact Us.](#)
DDBJ Read Annotation
Pipeline.
Development Team.

Select Query Files → Select Tools → Set QuerySet → Set Options → Submit

Running Status

Run Confirmation

Destination of mail
When the request is completed, the system sends an email to this address:
yn@nig.ac.jp

RUN

連絡先いれたら
実行可能

でも今は
押さないで！

Reference Genome Map [bowtie2]

Query sets	PairedOrientation	RunAccession	RunAlias	RowLength	QualityScore1	QualityScore2
Query set1	single	SRR331219	control			

genome sets
TAIR10
• all.fa

Command Options
bowtie2

Set optional parameters of the single-end analysis

Step1) Convert reference sequence
bowtie2-build -f refgenome.fasta bt2-idx

Step2) Map
bowtie2 -q -p 4 -x bt2-idx -U query1.fastq(.fasta) -S out.sam -u
out.unmapped

Step3) Convert the read alignment to .BAM format
samtools view -bS -o out.bam out.sam

Step4) Detect DNA polymorphism

「RUN を押した」と思ってください

処理状況は
こちらから

The screenshot shows the DDBJ Pipeline interface. At the top, a flowchart indicates the process: Select Query Files → Select Tools → Set QuerySet → Set GenomeSet → Set Map Options → Confirmation. Below this, a 'Running Status' button is highlighted. The main area is titled 'Status - Mapping' and includes tabs for Mapping Job, de novo Assembly Job, and Preprocessing Job. The 'Mapping Job' tab is selected. A table lists jobs with columns: Order, ID, UserID, Submission accession, P/S, Status, Tool, Read #, Read length, Genome size, Detail, Start time, End time, and Elapsed time. One row for job ID 4915 is highlighted with a yellow background. The 'View' button in the 'Detail' column for this row is circled in red. A progress bar is visible for another job. On the left, a 'JOB STATUS' sidebar shows steps: step1. Preprocessing, step1. Mapping (circled in red), step1. de novo Assembly, and step2-All status. A 'HELP' sidebar includes links for Help, Tutorial, Contact Us, and Development Team.

ID	User ID	Submission accession	P/S	Status	Tool	Read #	Read length	Genome size	Detail	Start time	End time	Elapsed time
4915	guest	SRA044892 control	S	running	Bowtie2	5,925,048	---	121 M	View	2013-01-11 23:23:03	---	
4913	guest	SRA049447 9870	S	complete	Bowtie2	14,278,727	---	4 M	View	2013-01-11 18:21:55	01:04:44	
4910	--	ERA000092 1_dpf_assay1_1 1_dpf_assay1_2 1_dpf_assay1_3 1_dpf_assay2_1 and more...	P	running	Bowtie2	375,421,197	---	1,379 M		2013-01-11 16:19:14	---	
4906	--	-- A1_Unshu_Paire	P	running	bwa	110,759,316	---	299 M		2013-01-10 16:24:05	---	
4905	--	-- Ion_mt20mergedc	S	complete	Bowtie2	148,666	---	16,520		2013-01-10 15:58:39	00:05:36	
4904	--	-- Ion_mt20mergedc	S	complete	Bowtie2	148,666	---	16,831		2013-01-10 15:54:22	00:06:24	
4903	--	--	S	complete	Bowtie2	148,666	---	16,589		2013-01-10 16:00:46	00:05:36	

DNA Import
Preprocessing Start
step-1
Preprocessing
Mapping / de novo Assembly
step-2
Workflow
Genome (SNP/Short Indel)
RNA-seq (Tag count)
ChIP-seq
JOB STATUS
step1. Preprocessing
step1. Mapping
step1. de novo Assembly
step2-All status

Tool (Version)			
Bowtie2 (2.0.0)			
RunAccession or Filename	Download		
SRR331219	SRR331219.fastq.bz2		
Genome set			
TAIR10			
Chromosome			
all.fa			
Download modified queries			
• SRR331219.fastq.gz (Original size 1.5 GB)			
Download merged pileup file			
• merged.pileup.gz (Original size 1.6 GB)			
• merged.sam.gz (Original size 1.4 GB)			
Download wgs file			
• out_WGS.fasta.gz (Original size 67.7 MB)			
Position errors	Map ratio	Depth, Coverage	
PDF download	total query # : 5,925,048 mapped query # : 5,037,456 map ratio : 85.020 %	coverage : 29886366 / 119482012 * 100 = 25.013 depth : 384468141 / 29886366 = 12.864	
Time			
Wait time	Start time	End time	
0: 1:12	2013-01-11 23:23:03	2013-01-12 00:19:10	

実行結果
(昨晩夜中)

all.fa	Command	Start time	End time	Log1	Log2	Result	MD5
	bowtie2-build -f all.fa refgenome	2013-01-11 23:23:03	2013-01-11 23:25:26	View		Download(194.4 MB)	MD5
	bowtie2 -p 4 -q -x refgenome -U SRR331219.fastq -S out.map --un out.unmapped	2013-01-11 23:26:07	2013-01-11 23:38:02		View	Download(463.0 MB)	MD5
	samtools view -bS -o out.bam out.map	2013-01-11 23:40:56	2013-01-11 23:42:28		View	Download(488.9 MB)	MD5
	samtools sort out.bam out2	2013-01-11 23:42:59	2013-01-11 23:44:52		View	Download(357.5 MB)	MD5

マッピング結果を眺めるには (BAM/SAM)

1. Tablet (<http://bioinf.scri.ac.uk/tablet/>)



Tablet

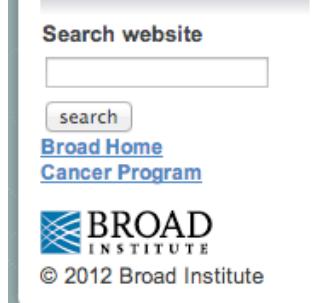
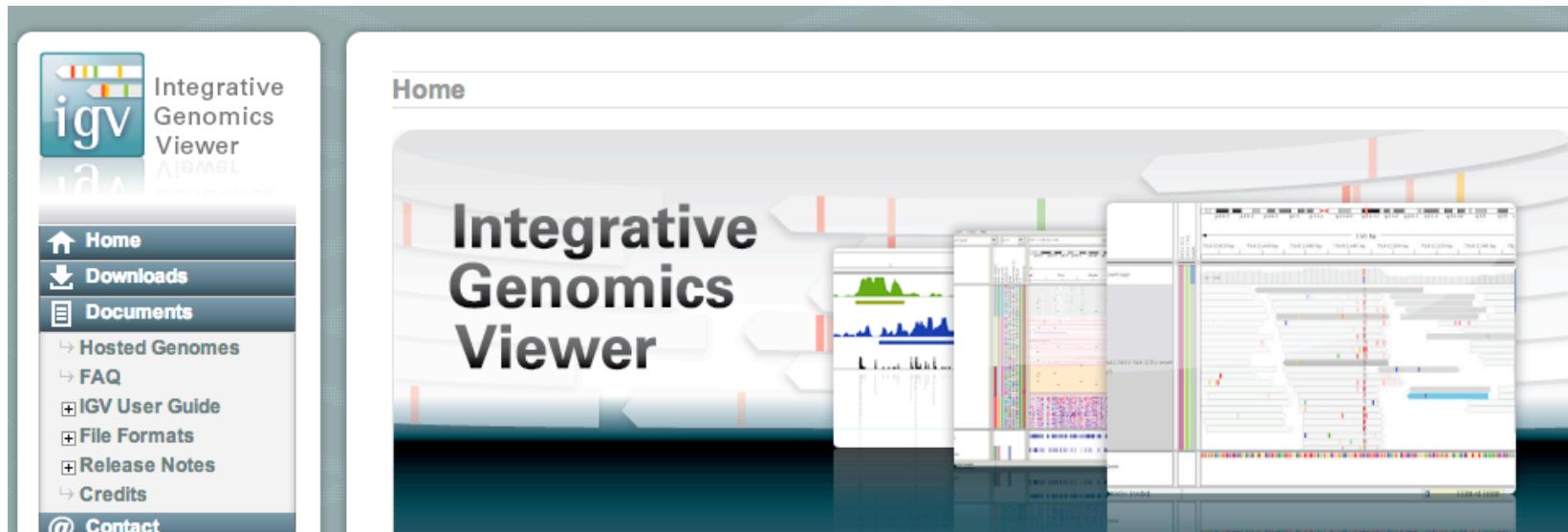
- [Tablet Homepage](#)
- [Download Tablet](#)
- [Screenshots](#)
- [Tablet FAQ](#)
- [Sample Data](#)
- [Assembly Conversion](#)
- [Papers and Presentations](#)
- [Privacy Policy](#)
- [Tablet World Map](#)
- [Online Help](#)

Our Software

- [CurlyWhirly](#)
- [Flapjack](#)
- [OPTIRas](#)
- [Strudel](#)
- [Tablet](#)
- [TetraploidMap](#)
- [TOPALi](#)

マッピング結果を眺めるには (BAM/SAM)

2. IGV (<http://www.broadinstitute.org/igv/>)



What's New

NEWS



December 18, 2012. IGV 2.2 has been released. See the [release notes](#) for more details.

April 20, 2012.

IGV 2.1 has been released. See the [release notes](#) for more details.

April 19, 2012. See our new [IGV paper](#) in *Briefings in Bioinformatics*.

Overview



The Integrative Genomics Viewer (IGV) is a high-performance visualization tool for interactive exploration of large, integrated genomic datasets. It supports a wide variety of data types, including array-based and next-generation sequence data, and genomic annotations.

Citing IGV

To cite your use of IGV in your publication:

Helga Thorvaldsdóttir, James T. Robinson, Jill P. Mesirov. [Integrative Genomics Viewer \(IGV\): high-performance genomics data visualization and exploration](#). *Briefings in Bioinformatics* 2012.

James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov. [Integrative Genomics Viewer](#). *Nature Biotechnology* 29, 24–26 (2011).

Funding

Development of IGV is made possible by funding from the [National Cancer Institute](#), the [National Institute of General Medical Sciences](#) of the [National Institutes of Health](#), and the [Stanford University](#).

【再掲】スパコン利用申請はこちちらです

- <http://www.ddbj.nig.ac.jp/system/supercom/supercom-apl.html>
- [DDBJ スーパーコンピュータ] で検索

The screenshot shows the DDBJ (DNA Data Bank of Japan) website. The header features the DDBJ logo, the text "DNA Data Bank of Japan", and links for ENGLISH, Twitter, and RSS. Below the header is a navigation bar with links for HOME, 塩基配列の登録, 利用の手引き (highlighted in yellow), 検索・解析, FTP・WebAPI, レポート・統計, お問い合わせ, and サイト内検索. The main content area is titled "Supercomputer System Application". It contains text about logging in to the supercomputer system, mentioning the "SAKURA" system, the "MSS" large-scale registration system, data correction and update, and the "DDBJ Sequence Read Archive" and "DDBJ Trace Archive". It also mentions the "BioProject Database". A note at the bottom states: "For those who hope to log in to the supercomputer, please apply from here. After application submission, we will send a registration certificate by mail." The URL shown in the browser is <http://www.ddbj.nig.ac.jp/system/supercom/supercom-apl.html>.