



遺伝研大量研・DDBJセンター

中村保一

# DRA

DDBJ Sequence Read Archive

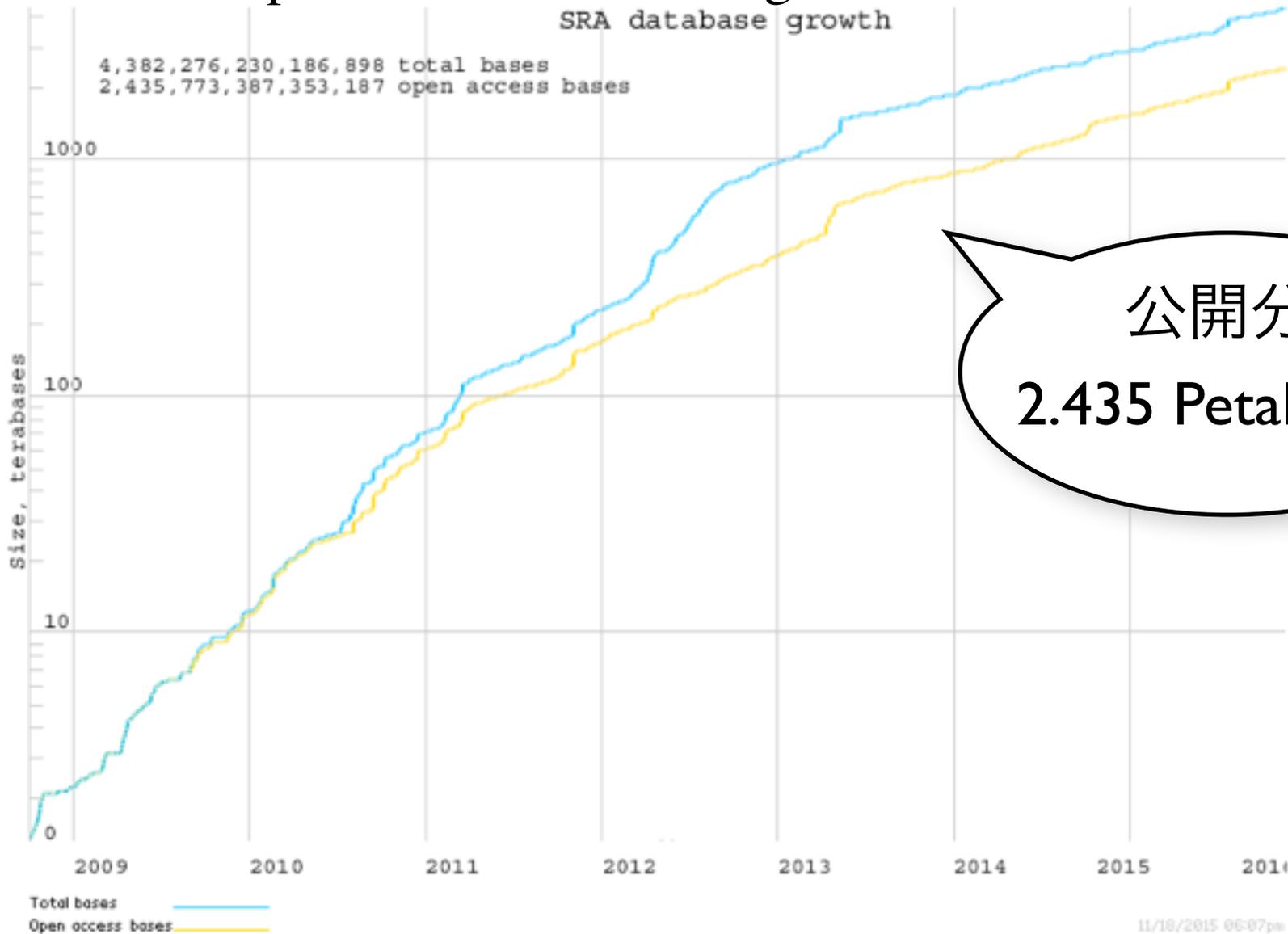
# DDBJ Sequence Read Archive (DRA)

新世代シーケンサから出力される配列や  
アライメントデータを登録・公開



# SRA growth (NCBI)

<http://trace.ncbi.nlm.nih.gov/Traces/sra>



# DRAウェブサイト ⇒ [DRA] で検索

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

登録関係情報



## Sequence Read Archive

Login & Submit | Databases ▾ | English | Contact

Google™ カスタム検索



Home

Submission ▾

Search

Download ▾

Pipeline

About

解析パイプライン

DDBJ Sequence Read Archive (DRA) は Illumina 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を蓄積・提供する。Nucleotide Sequence Database Collaboration (INSDC) のメンバーであり、INSDC の Sequence Read Archive (ERA) との国際協力のもと、運営されています。従来のキャピラリーシーケンサからの出力データを蓄積・提供する DRA Archive にご登録ください。

データ検索

データ取得



検索

データをキーワード、生物名、シーケンサなどで検索する



登録

新型シーケンサからの生データやアライメントデータを登録する



動画マニュアル

DRA の利用方法や登録方法を解説している動画を見る

# 公開データの DRA Search での検索

公開データは EBI SRA / NCBI SRA と共有されています



The screenshot shows the DRA Search interface. At the top, there are search filters for Accession (DRA000001), Organism (Abiotrophia defectiva ATCC 49176), StudyType (Epigenetics), and Platform (ILLUMINA). Below the filters, there are buttons for 'Search' and 'Clear'. A callout bubble points to the search filters with the text '生物名 etc での絞り込み' (Filtering by organism name, etc.).

The main content area is divided into two sections. On the left, there is a 'Statistics' section with a table of 'Released Entries' and a table of 'Organism' counts. On the right, there is a 'Search Results (358 studies)' section with a table of search results. A callout bubble points to this section with the text '検索結果リスト' (Search results list).

The search results table has columns: #, STUDY, SUBMISSION, STUDY\_TITLE, STUDY\_TYPE, ORGANISM, and CENTER\_NAME. The first row is highlighted in orange and labeled 'DRP000003'. A callout bubble points to this row with the text 'ダウンロード' (Download).

Below the search results, there is a 'Study Detail' section for the selected study (DRP000003). It includes fields for Title, Abstract, Description, Project ID, and Center Name. A callout bubble points to this section with the text '詳細 (メタデータ記述)' (Details (Metadata description)).

On the right side of the study detail, there is a 'Navigation' section with links for Submission, Experiment, and Sample, each with a download icon. A callout bubble points to this section with the text 'ダウンロード' (Download).

# 解析パイプラインも提供しています

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



## Sequence Read Archive

Home

Submission ▾

Search

Download ▾

Pipeline

About

解析パイプライン

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。従来のキャピラリー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。



### 検索

データをキーワード、生物名、シーケンサなどで検索する



### 登録

新型シーケンサからの生データやアライメントデータを登録する



### 動画マニュアル

DRA の利用方法や登録方法を解説している動画を見る

# DDBJ 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

**DDBJ**  
DNA Data Bank of Japan

**ACCOUNT**  
login ID [yaskaz]  
Logout  
Change password

**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.



## 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	
<input type="checkbox"/>	<a href="#">BLAT</a>	<a href="#">Help</a>	34	✓						✓					Single-end analysis only
<input type="checkbox"/>	<a href="#">Maq</a>	<a href="#">Help</a>	0.7.1	✓		✓				✓	✓	✓	✓	✓	
<input type="checkbox"/>	<a href="#">bwa</a>	<a href="#">Help</a>	0.5.9	✓		✓				✓				✓	
<input type="checkbox"/>	<a href="#">SOAP</a>	<a href="#">Help</a>	2.21	✓		✓				✓	✓			✓	
<input type="checkbox"/>	<a href="#">Bowtie</a>	<a href="#">Help</a>	0.12.7	✓	✓	✓				✓	✓			✓	
<input type="checkbox"/>	<a href="#">TopHat</a>	<a href="#">Help</a>	1.0.11	✓		✓				✓				✓	
<input type="checkbox"/>	<a href="#">Bowtie2</a>	<a href="#">Help</a>	2.0.0	✓	✓	✓				✓	✓			✓	For reads longer than about 50 bp, Bowtie2 is generally faster, more sensitive, and uses less memory than Bowtie1.

### de novo Assembly

Total limit = 22 Gbp

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

# DDBJ 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

The screenshot displays the DDBJ pipeline interface. At the top, a workflow diagram shows the steps: Select Query Files → Select Tools → Set QuerySet → Running Status. The main content area is titled 'Specifying Database of Reference' and features a 'Major genome sets' section. This section includes a dropdown for 'Organisms' (currently showing 'Arabidopsis thaliana') and a 'Genome sets' dropdown (currently showing 'TAIR8'). A red arrow points to the 'Arabidopsis thaliana' selection in the 'Organisms' dropdown. Below the 'Genome sets' dropdown, there are checkboxes for 'all.fa' and individual chromosomes (chr01.fa to chr17.fa). The left sidebar contains sections for 'ACCOUNT' (login ID, Logout, Change password), 'ANALYSIS' (Data setup, DRA Start, FTP upload), 'JOB STATUS' (step1. Preprocessing, step1. Mapping, step1. de novo Assembly, step2-All status), and 'HELP'. The right side of the image shows two zoomed-in views of the 'Major genome sets' dropdown menu. The top view shows the 'Oryza sativa japonica' dropdown with a list of genome sets including 'IRGSP Releases Build 4.0', 'IRGSP Releases Build 5.0', and 'tigr version5.0'. The bottom view shows the 'Homo sapiens' dropdown with a list of genome sets including 'Homo sapiens Feb. 2009 (hg19)', 'Mar.2006 (hg18)', and 'May.2004 (hg17)'.

# DDBJ パイプライン、体験してみましよう

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



## DDBJ Read Annotation Pipeline

English

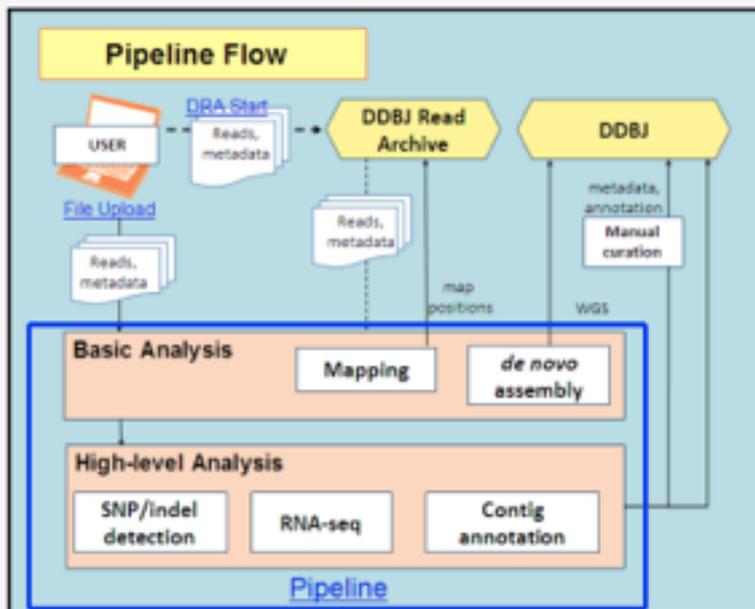
Japanese

DDBJ Read Annotation Pipeline is a cloud-computing based analytical platform for next-generation sequencing data.

LOGIN

New account

Login as "guest"



ゲストとして  
ログイン

User ID:

or

[Check current jobs](#)

\* by the guest account.

### Manual & tutorial

- [Japanese Tutorial](#)
- [English manual](#)
- [DBCLS togotv Tutorial video 1 \(JP\) - Reference Genome Mapping](#)
- [DBCLS togotv Tutorial video 2 \(JP\) - De novo Assembly](#)
- [FAQ : How to upload and register query files to DDBJ Pipeline \(JP\)](#)
- [Tutorial : How to run HGAP for PacBio sequence read on DDBJ Pipeline \(JP\)](#)

Tweets

Follow

pipeline

11 Jun

# 処理に使うNGSの配列ファイルの用意



## 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.

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	DRA000001.submission.xml	Download	View
Sample	DRS000001	DRS000001	DRA000001.sample.xml	Download	View
Study	DRP000001	DRP000001	DRA000001.study.xml	Download	View
Experiment	DRX000001	DRX000001	DRA000001.experiment.xml	Download	View
Run	DRR000001	DRR000001	DRA000001.run.xml	Download	View

**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
<input type="checkbox"/>	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



# 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

**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 → Set Map Options → Confirmation

Running Status

Selecting

NEXT

FTP upload Private DRA entry Import public DRA Preprocessing HTTP upload

List of your uploaded files by FTP client. [Add new files](#)

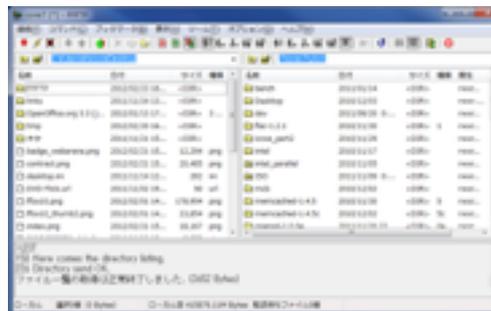
Filename	Description	Layout	Instrument model	File size
<input type="checkbox"/> demo.fastq	demo study	single	ILLUMINA	340.9 KB
<input type="checkbox"/> coli_allmands_5runs.fastq	coliAll	single	LS454	139.9 MB
<input type="checkbox"/> 56-Rb1_S6_L001_R1_001.fastq.gz.0 (more 1 files)	150220 Rb1	paired	ILLUMINA	304.2 MB
<input type="checkbox"/> 56-Rb1_S6_L001_R1_001.fastq.gz.0 (more 1 files)	150220 Rb1	paired	ILLUMINA	304.2 MB
<input type="checkbox"/> demo.fastq.6	DEMO	single	ILLUMINA	340.9 KB
<input type="checkbox"/> 1.fas	lab	single	ILLUMINA	0 byte
<input type="checkbox"/> 1.fas	RNR	single	ILLUMINA	0 byte
<input type="checkbox"/> RRRRR.fna	RNR-LAB	single	ABI_SOLID	12.3 MB
<input type="checkbox"/> AC_R1.fastq (more 1 files)	Assembly of Honey Bee Mitogenome	paired	ILLUMINA	188.2 MB
<input type="checkbox"/> C66BHACXX_PG0614_98A0501_H1_L007_R1-50bp_QV10.fastq	test_1	single	ILLUMINA	337.6 MB
<input type="checkbox"/> demo.fastq	ddbj pipeline demo	single	ILLUMINA	340.9 KB
<input type="checkbox"/> demo.fastq	ddbj pipeline demo	single	ILLUMINA	340.9 KB

DELETE NEXT

# FTP (file transfer protocol) クライアント



Cyberduck



FFFTP

```
iMac1- yn$ ftp yanakamu@gw.ddbj.nig.ac.jp
yanakamu@gw.ddbj.nig.ac.jp's password:
```

Cyberduck のような専用クライアントを使えば  
ドラッグ&ドロップでファイル転送可能。コマン  
ドでの送受信もたいして難しくありませんが。

# 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.



Running Status

## Selecting Query Files

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

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

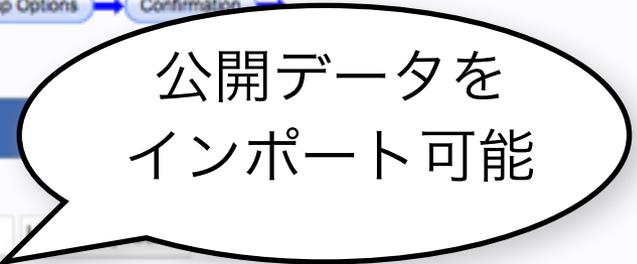
Accession Number can find here.  
[DRA Search](#)

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	SRA045657	2015-11-18 15:56:10.735
✘ preparing	SRA221099	2015-10-28 17:33:00.955
✔ done	SRA073646	2015-10-22 15:34:33.954
✔ done	ERA000035	2015-10-21 16:43:31.226
✘ preparing	SRA176628	2015-10-18 23:11:12.59
✔ done	SRA012701	2015-09-28 12:37:52.009
✘ preparing	SRA245648	2015-07-29 14:52:11.178
✔ done	SRA010042	2015-07-24 11:39:14.19
✔ done	ERA209804	2015-07-24 10:56:30.546
✔ done	SRA009815	2015-07-20 20:51:21.762



# 今回はupload済のエントリから



Running Status

## Selecting Query Files

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	DRA000001.submission.xml	Download	View
Sample	DRS000001	DRS000001	DRA000001.sample.xml	Download	View
Study	DRP000001	DRP000001	DRA000001.study.xml	Download	View
Experiment	DRX000001	DRX000001	DRA000001.experiment.xml	Download	View
Run	DRR000001	DRR000001	DRA000001.run.xml	Download	View

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
<input type="checkbox"/>	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

checkして

- 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 Teams.

**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.

**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	
<input type="checkbox"/>	BLAT ⓘ	◆	34	✓											Single-end analysis only
<input type="checkbox"/>	bwa ⓘ	◆	0.6.1	✓		✓	✓	✓	✓					✓	
<input type="checkbox"/>	Bowtie ⓘ	◆	0.12.7	✓	✓	✓	✓	✓	✓	✓				✓	
<input type="checkbox"/>	TopHat ⓘ	◆	1.0.11	✓		✓	✓	✓	✓					✓	
<input type="checkbox"/>	Bowtie2 ⓘ	◆	2.0.0	✓	✓	✓	✓	✓	✓	✓				✓	For reads longer than about 50 bp, Bowtie2 is generally faster, more sensitive, and uses less memory than Bowtie1.
<input type="checkbox"/>	TopHat2 ⓘ	◆	2.0.9	✓		✓	✓	✓	✓					✓	

**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 checked="" type="checkbox"/>	Velvet ⓘ	◆	1.2.10	✓		✓	✓	We severe recommend when performing Velvet, total length of those reads is up to 22G bp Maximum K-mer value is 64.
<input type="checkbox"/>	Trinity ⓘ	◆	r2013-02-25	✓		✓		RNA-Seq De novo Assembly
<input type="checkbox"/>	Platanus ⓘ	◆	1.2.2	✓		✓		
<input type="checkbox"/>	HGAP ⓘ	◆	Protocol3 (v 2.2.0)					HGAP Pipeline for PacBio Sequence based on SMRT Analysis v2.2.0. For bax.h5 file only. (Beta version)

Mapping Contigs by de novo Assemble to Reference Sequences.  
The contigs will be aligned to reference genome.

Tool	Comment
<input type="radio"/> BLAT	Single-end analysis only

BACK **NEXT**



# 配列とペア形式のセットを選択



## 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

RESET BACK NEXT

**Paired-end analysis**

Layout of paired sequence. 5'-3' 3'-5'

Run	ACCESSION	Read length	Quality Score
<input checked="" type="checkbox"/>	D13000001 -><-	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



## 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

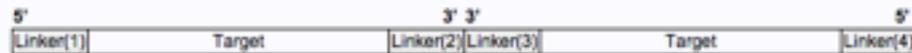
Contact Us.  
DDBJ Read Annotation  
Pipeline.  
Development Team.

## Generating Query Sets from Query Read Files

RESET BACK NEXT

### Paired-end analysis

Layout of paired sequence. 5'-3' 3'-5'



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びます



- 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

## Setting for De Novo Assembly

BACK NEXT

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.  
Velvet 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

# 連絡先をいれたら、RUN ボタン押すだけ！

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

## Run Confirmation

BACK

Destination of mail

When the request is completed, the system sends an email to this address.

\* Required

Maximum time is limited 60 days after submission.

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.  
Velvet output\_directory/ 23 -fastq -shortPaired shuffle\_query\_pe.fastq

Step2) Assembly

Velvetg output\_directory/ -ins\_length 300 -exp\_cov auto

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 100 out\_WGS.txt

BACK

連絡先いれたら  
RUN可能

guestは  
runできません

# 「RUN を押した」と思ってください



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

## HELP

HELP ⓘ

TUTORIAL

Contact Us  
DDBJ Read Assembly  
Pipeline  
Development Team

## Status - de novo Assembly

Mapping Job

de novo Assembly Job

Preprocessing Job

### Order

Sort by: ID Descending

Show Only Your Own Job

Reload

Delete \*

page 1 NEXT >

	ID	UserID	Submission accession	P/S	Status	Tool	Read #	Read length	Assembly detail	Mapping detail	Start time End time	Elapsed time
<input type="checkbox"/>	20023	---	SRA020075 CGAa	S	complete	SOAPdenovo	866,837	---			2015-11-18 18:49:44 2015-11-18 18:59:58	00:10:14
<input type="checkbox"/>	20014	---	SK1_R1 by Preq	P	complete	SOAPdenovo	3,744,114	---			2015-11-18 10:36:14 2015-11-18 11:54:48	01:18:34
<input type="checkbox"/>	20010	---	SK1_R1 by Preq	P	complete	Velvet	3,744,114	---			2015-11-18 08:39:52 2015-11-18 09:15:58	00:36:06
<input type="checkbox"/>	20008	---	Aster-s1 Aster-s2	P	running	Trinity	100,829,136	---			2015-11-17 23:18:12 ---	
<input type="checkbox"/>	20007	---	Aster-r1 Aster-r2	P	running	Trinity	91,276,607	---			2015-11-17 23:09:00 ---	
<input type="checkbox"/>	20006	---	As-IS	P	complete	Trinity	46,700,225	---			2015-11-17 21:07:32 2015-11-18 17:46:39	20:39:07
<input type="checkbox"/>	20005	---	As-IS	P	complete	Trinity	28,101,597	---			2015-11-17 21:03:17 2015-11-18 07:46:04	10:42:47
<input type="checkbox"/>	20003	---	As-IR	P	complete	Trinity	18,598,628	---			2015-11-17 21:00:52 2015-11-18 07:46:04	06:57:37

処理状況は  
こちらから

## 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.

DCBJ 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	<a href="#">DRR000001.fastq.gz</a> <a href="#">DRR000001_1.fastq.gz</a> <a href="#">DRR000001_2.fastq.bz2</a>	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

### 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	<a href="#">View</a>	<a href="#">View</a>	<a href="#">Download(174.7 MB)</a>	<a href="#">MD5</a>

BACK

アセンブル結果の  
基本情報

結果ファイル

# Mappingの例 (DRASearch+pipeline)

Accession :   
Organism :   
CenterName :   
Keyword : alternative splicing Arabidopsis

シロイヌナズナ  
alternative splicing

Show 20 records Sort by Study Search Clear

## Search Results ( 3656 records )

<< < 1 / 183 Page > >>

Filtered by  
document type:study(1890) sample(1020) experiment(600) submission(103) run(41) analysis(2)  
organism:Arabidopsis thaliana(1704) Homo sapiens(523) Mus musculus(462) Drosophila melanogaster(55) Saccharomyces cerevisiae(31)  
Arabidopsis lyrata(30)

#	META_FILE	ACCESSION	STUDY	STUDY_TITLE	STUDY_TYPE	ORGANISM	BASES	SUBMITTED	CENTER_NAME
1	<a href="#">SRA050132.study.xml</a> </?xml version="1.0" encoding="UTF-8"?> <STUDY center_name="NCSU" alias="Arabidopsis - Pseudomonas alternative splicing study" accession="SRP010938" />	SRP010938	SRP010938	Arabidopsis thaliana strain:Columbia (Col-0) Transcriptome or Gene expression	Transcriptome Analysis	Arabidopsis thaliana	85.2G	2012-02-15	NCSU
2	<a href="#">SRA009031.study.xml</a> </SUBMITTER_ID> </IDENTIFIERS> <DESCRIPTOR> <STUDY_TITLE>Transcriptome-wide map of alternative splicing in Arabidopsis	SRP000935	SRP000935	Transcriptome-wide map of alternative splicing in Arabidopsis	Transcriptome Analysis	Arabidopsis thaliana	12.5G	2009-07-07	OSU-CGRB
3	<a href="#">SRA050132.submission.xml</a> <?xml version="1.0" encoding="UTF-8"?> <SUBMISSION alias="Arabidopsis alternative splicing project" />	SRA050132	SRP010938	Arabidopsis thaliana strain:Columbia (Col-0) Transcriptome or Gene expression	Transcriptome Analysis	Arabidopsis thaliana	85.2G	2012-02-15	NCSU
4	<a href="#">SRA198305.submission.xml</a> splicing-contributed proteome diversity in Arabidopsis thaliana center_name="Institute of Genetics"	SRA198305	SRP049534	Transcriptome survey of alternative splicing-contributed proteome diversity in Arabidopsis thaliana	Transcriptome Analysis	Arabidopsis thaliana	84.3G		BioProject
5	<a href="#">SRA044892.study.xml</a> </?xml version="1.0" encoding="UTF-8"?> <STUDY center_name="Institute of Plant and Microbial Biology, Academia" alias="Alternative Splicing" />	SRP007763	SRP007763	Genome-wide detection of context-sensitive alternative splicing in Arabidopsis roots	Other	Arabidopsis thaliana	3.6G		Institute of Plant and Microbial Biology, Academia
6	<a href="#">DRA002223.run.xml</a> :01:00+09:00"> <TITLE>Phytochrome controls alternative splicing to mediate light responses in Arabidopsis	DRR018422 DRR018423 DRR018425 DRR018424 DRR018428 DRR018427 DRR018426	DRP002301			Arabidopsis thaliana	202.4G	2014-04-15	

# データのIDはこちら

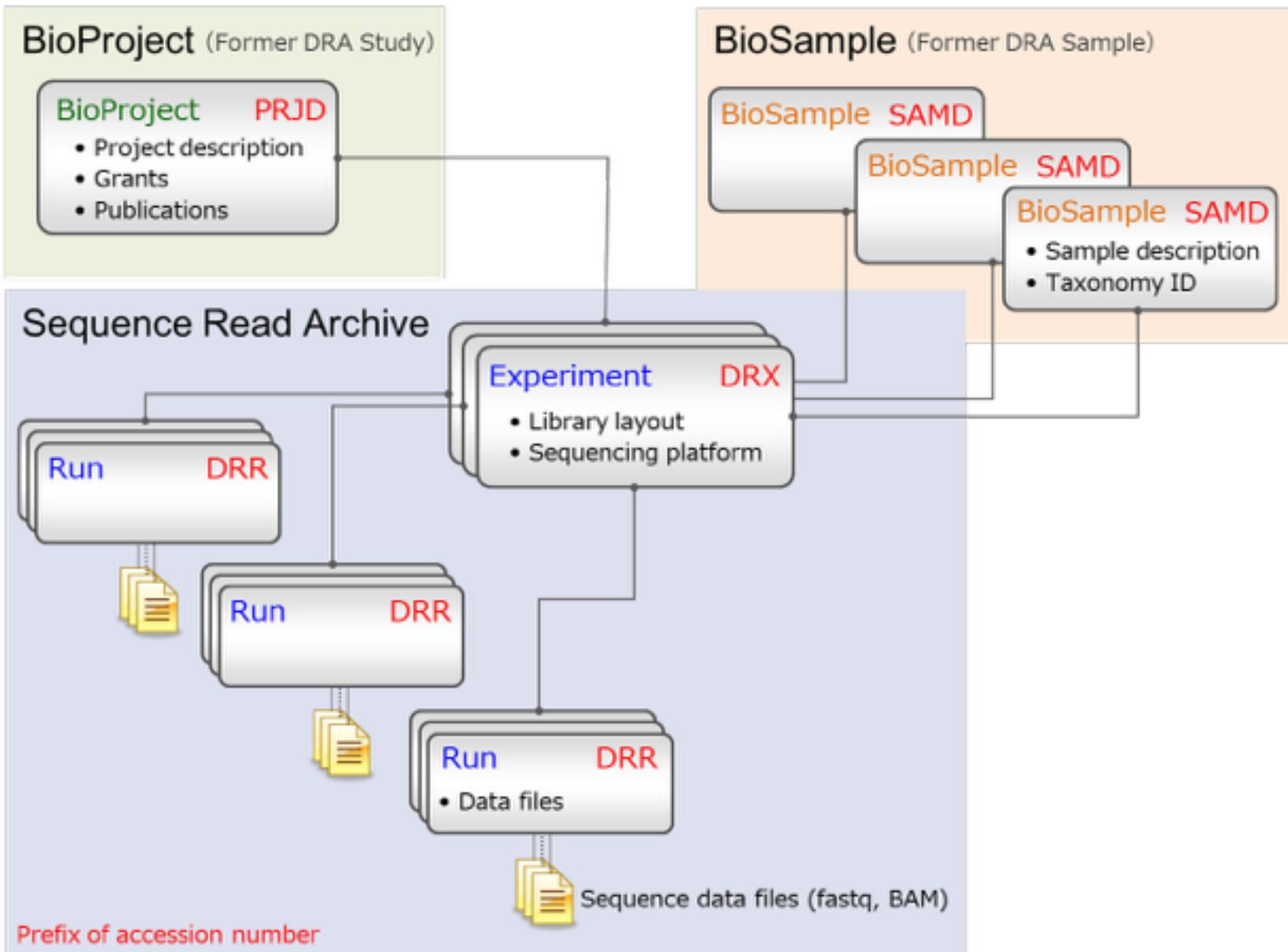
SRP007763

Study Detail	
Title	Genome-wide detection of context-sensitive alternative splicing in Arabidopsis roots
Study Type	Other
Abstract	To analyze context-sensitive changes in pre-mRNA splicing pattern and gene expression, we mapped the transcriptome of iron-deficient and iron-sufficient Arabidopsis roots using the RNA-seq technology. RNA-seq data were analyzed with a newly developed software package, RACKJ (Read Analysis & Comparis .. <a href="#">[more]</a> )
Description	
Center Name	Institute of Plant and Microbial Biology, Academia

Navigation	
Submission	<a href="#">SRA044892</a> <a href="#">FTP</a>
	<a href="#">SRA045009</a> <a href="#">FTP</a>
Experiment	<a href="#">SRX092048</a> <a href="#">FASTQ</a> <a href="#">SRA</a>
	<a href="#">SRX092049</a> <a href="#">FASTQ</a> <a href="#">SRA</a>
	<a href="#">SRX092050</a> <a href="#">FASTQ</a> <a href="#">SRA</a>
Sample	<a href="#">SRS256250</a>
	<a href="#">SRS257585</a>
	<a href="#">SRS257586</a>
	<a href="#">SRS257587</a>

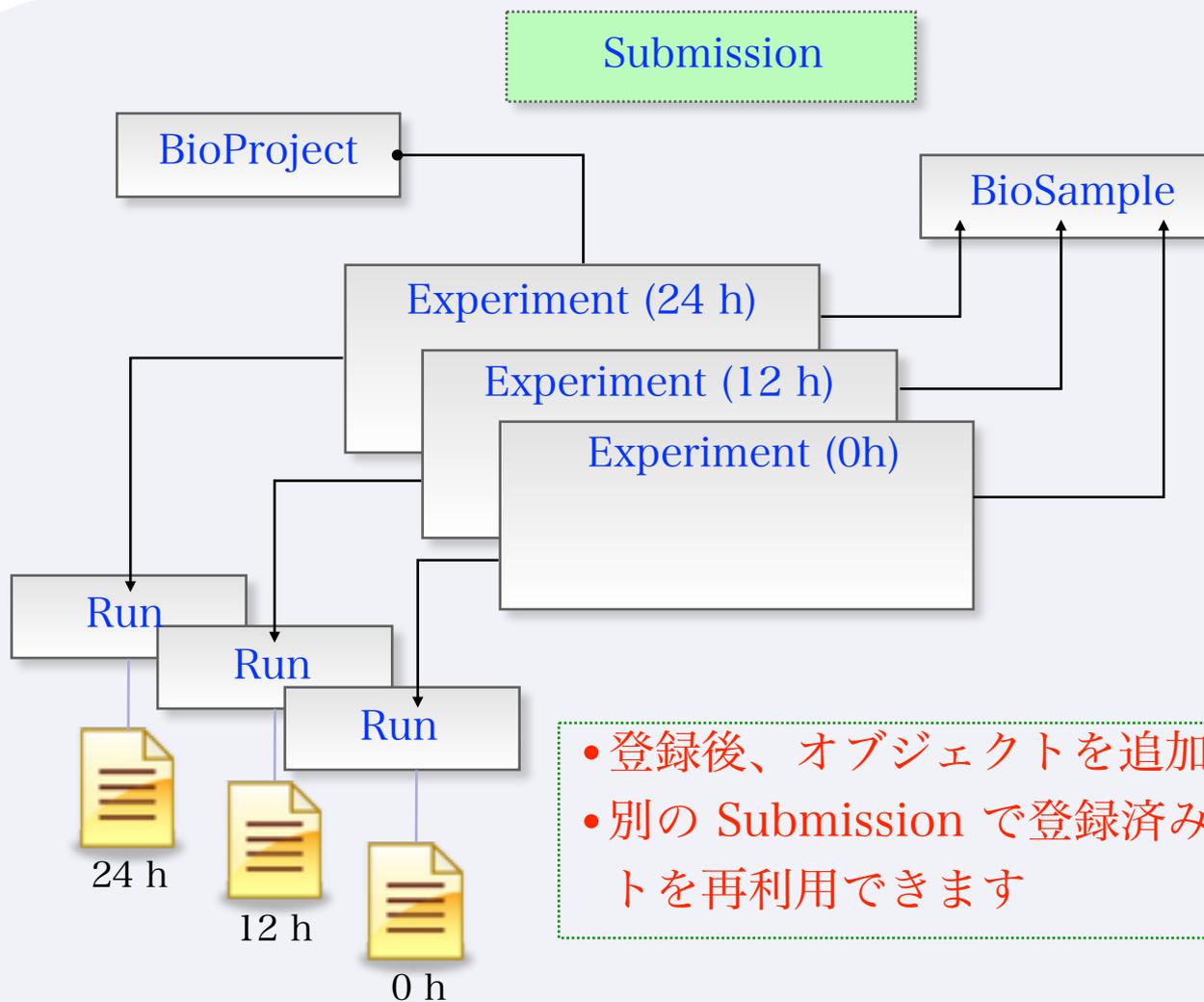
# SRA の Metadata

<https://trace.ddbj.nig.ac.jp/dra/submission.html>



# 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



- 登録後、オブジェクトを追加できます
- 別の Submission で登録済みのオブジェクトを再利用できます

# p.ddbj.nig.ac.jp を開き、さっきのIDを入力

**DDBJ**  
DNA Data Bank of Japan

ACCOUNT  
login ID [guest]  
Logout

ANALYSIS  
Data setup  
DRA Start  
FTP upload  
DRA Import

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

JOB STATUS  
step1. Preprocessing  
step1. Mapping / 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

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	SRA045657	2015-11-18 15:58:10.735
✘ preparing	SRA221099	2015-10-28 17:33:00.955
✔ done	SRA073646	2015-10-22 15:34:33.954
✔ done	ERA000035	2015-10-21 16:43:31.226
✘ preparing	SRA176628	2015-10-18 23:11:12.59
✔ done	SRA012701	2015-09-28 12:37:52.009
✘ preparing	SRA245648	2015-07-29 14:52:11.178
✔ done	SRA010042	2015-07-24 11:39:14.19
✔ done	ERA209804	2015-07-24 10:56:30.546
✔ done	SRA009815	2015-07-20 20:51:21.762

でも今は  
押さないで！

# SRA044892 をロードしておきました



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
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** Other

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.)

DELETE NEXT

# Bowtie2 を選んで NEXT



Running Status

## Selecting Tools for Basic Analysis of DDBJ ANNOTATION PIPELINE

BACK **NEXT**

### 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"/>	<a href="#">BLAT</a>		34	✓					✓						Single-end analysis only
<input type="checkbox"/>	<a href="#">bwa</a>		0.6.1	✓		✓	✓	✓	✓					✓	
<input type="checkbox"/>	<a href="#">Bowtie</a>		0.12.7	✓	✓	✓	✓	✓	✓	✓				✓	
<input type="checkbox"/>	<a href="#">TopHat</a>		1.0.11	✓		✓	✓	✓	✓					✓	
<input checked="" type="checkbox"/>	<a href="#">Bowtie2</a>		2.0.0	✓	✓	✓	✓	✓	✓	✓				✓	For reads longer than about 50 bp, Bowtie2 is generally faster, more sensitive, and uses less memory than Bowtie1.
<input type="checkbox"/>	<a href="#">Bowtie</a>		2.0.9	✓		✓	✓	✓	✓	✓				✓	

### de novo Assembly

Total limit = 22 Gbp

	Tool	Help	Version	Base space	Color space	Paired-end	MSS(WGS)	Comment
<input type="checkbox"/>	<a href="#">SOAPdenovo</a>		1.05	✓		✓		
<input type="checkbox"/>	<a href="#">ABySS</a>		1.3.2	✓		✓		Maximum K-mer value is 64.
<input type="checkbox"/>	<a href="#">Velvet</a>		1.2.10	✓		✓	✓	We severe recommend when performing Velvet, total length of those reads is up to 22G bp. Maximum K-mer value is 64.
<input type="checkbox"/>	<a href="#">Trinity</a>		r2013-02-25	✓		✓		RNA-Seq De novo Assembly
<input type="checkbox"/>	<a href="#">Platanus</a>		1.2.2	✓		✓		
<input type="checkbox"/>	<a href="#">HGAP</a>		Protocol3 (v 2.2.0)					HGAP Pipeline for PacBio Sequence based on SMRT Analysis v2.2.0. For bax.h5 file only. (Beta version)

### Mapping Contin by de novo Assemble to Reference Sequences

**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.

# 配列を選んで confirm, NEXT



Running Status

## Generating Query Sets from Query Read Files

RESET BACK NEXT

Single analysis

Layout of single sequence.

5'				3'
	Linker(1)	Target	Linker(2)	

Run ID	Read length	Quality Score
<input checked="" type="checkbox"/> SRR331219 ->		bp
<input checked="" type="checkbox"/> SRR331224 ->		bp

confirm

QUERY SET

RESET BACK 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.  
Development Team.

# リファレンスに TAIR10 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.



Running Status

## Specifying Database of Reference Genome

RESET BACK NEXT

### Major genome sets

Organisms

Genome sets

- all.fa
- 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



Running Status

## Setting for Reference Genome Mapping

BACK **NEXT**

**bowtie2**

### Set optional parameters of the single-end analysis

Step1) Convert reference sequence

```
bowtie2-build [input] -f refgenome.fasta bt2-idx
```

Step2) Map

```
bowtie2 -q -p 4 [input] -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.

<input type="radio"/>	samtools pileup -c [input] -c -f refgenome.fasta out.bam   bcftools view
<input checked="" type="radio"/>	samtools mpileup -u -C50 -BQ0 -d10000000 [input] -f refgenome.fasta out.bam   bcftools view -bvcg [input] -> out.var.raw.bcf
<input type="radio"/>	bcftools view out.var.raw.bcf   vcftools.pl varFilter -D10000 > out.var.ft.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](#).

perl getConsGeno\_4pipeline.pl pileupFile [input] Not to include insertion of pileupped reads.  out\_WGS.bt

*\* Threshold of insertion of pileupped reads: the quality threshold for indels <= 50 and allele constitutes 80% of pileupped reads.*

BACK 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. Development Team.

# 終了したらメールが来ます



Running Status

## Run Confirmation

BACK

Destination of mail  
When the request is completed, the system sends an email to this address.  
 \* Required

Reference Genome Map [bowtie2]

Query sets					
Query set1					
PairedOrientation	RunAccession	RunAlias	RowLength	QualityScore1	QualityScore2
single	SRR331219	control			
single	SRR331224	iron			

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

Please choose one of the following.

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

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

連絡先いれたら  
RUN可能

guestは  
runできません

- 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.

# 「RUN を押した」と思ってください



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

Running Status

## Status - Mapping

Mapping Job    de novo Assembly Job    Preprocessing Job

Order

Sort by: ID    Descending     Show Only Your Own Job    Reload

Delete \*

page 1    NEXT >

	ID	UserID	Submission accession	P/S	Status	Tool	Read #	Read length	Genome size	Detail	Start time End time	Elapsed time
<input type="checkbox"/>	5499	guest	DRA000001 DRR000001	P	complete	SOAP	9,977,388	36	121 M	<a href="#">View</a>	2013-04-26 15:38:54	00:18:45
<input type="checkbox"/>	5267	guest	DRA000001 DRR000001	P	complete	Bowtie2	9,977,388	36	121 M	<a href="#">View</a>	2013-04-08 10:30:52	01:30:24
<input type="checkbox"/>	5235	guest	---	S	error	BLAT	---	---	3,197 M	<a href="#">View</a>	---	---
<input type="checkbox"/>	5167	guest	DRA000001 DRR000001	P	complete	Maq	9,977,388	36	3,197 M	<a href="#">View</a>	2013-03-14 11:31:11	00:31:26
<input type="checkbox"/>	5137	guest	DRA000001 DRR000001	P	complete	TopHat	9,977,388	36	253 M	<a href="#">View</a>	2013-03-05 09:53:55	00:46:30
<input type="checkbox"/>	5066	guest	ERA000095 ID49_020708_2	P	complete	Bowtie	5,925,048	---	121 M	<a href="#">View</a>	2013-03-05 10:40:25	00:47:03
<input type="checkbox"/>	5055	guest	ERA000095 ID49_020708_2	P	complete	bwa	3,543,021	36	12 M	<a href="#">View</a>	2013-02-15 12:26:14	00:47:42
<input type="checkbox"/>	5050	guest	ERA000095	P	complete	bwa	3,543,021	36	6 M	<a href="#">View</a>	2013-02-15 13:13:17	00:26:48
<input type="checkbox"/>	5055	guest	ERA000095 ID49_020708_2	P	complete	bwa	3,543,021	36	12 M	<a href="#">View</a>	2013-02-07 20:33:31	00:47:42
<input type="checkbox"/>	5050	guest	ERA000095	P	complete	bwa	3,543,021	36	6 M	<a href="#">View</a>	2013-02-07 21:21:14	00:39:30

処理状況は  
こちらから

### 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

de novo Assembly

step2-All status

### HELP

HELP ?

### TUTORIAL

Contact Us.  
DDBJ Read Annotation  
Pipeline.  
Development Teams.

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

Tool (Version)  
Bowtie2 (2.0.0)

RunAccession or Filename	Download	Read length	Alias
SRR331219	<a href="#">SRR331219.fastq.bz2</a>	N.A. bp	control

**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
<a href="#">PDF download</a>	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

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	<a href="#">View</a>		<a href="#">Download(194.4 MB)</a>	<a href="#">MD5</a>
	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		<a href="#">View</a>	<a href="#">Download(463.0 MB)</a>	<a href="#">MD5</a>
	samtools view -bS -o out.bam out.map	2013-01-11 23:40:56	2013-01-11 23:42:28		<a href="#">View</a>	<a href="#">Download(488.9 MB)</a>	<a href="#">MD5</a>
	samtools sort out.bam out2	2013-01-11 23:42:59	2013-01-11 23:44:52		<a href="#">View</a>	<a href="#">Download(357.5 MB)</a>	<a href="#">MD5</a>

実行結果

Development Team.

現在地: Home

2015年11月19日

- Language/言語
- ホーム
- このサイトへのログイン
- Login (スパコンユーザでログイン可)
- システム構成
  - ハードウェア構成
  - ソフトウェア構成
  - プログラミング環境
  - 利用可能バイオツール
  - 利用可能OSS
  - 利用可能DB

### 重要なお知らせ

- 公開日 表題
- 2015年11月12日 【定期メンテナンス2】 12月11日～12月16日 国立遺伝学研究所法定停電に伴うサービス停止のお知らせ
- 2015年10月30日 【UGL障害】 スーパーコンピュータシステム UGL動作不良のお知らせ
- 2014年9月10日 【スパコンユーズ会】 会議報告
- 2014年3月4日 2014年3月5日からのスパコンPhase2システムご利用方法について

### 国立遺伝学研究所 スーパーコンピュータシステム(NIG SUPERCOMPUTER)とは

大学共同利用機関法人 情報システム研究機構 国立遺伝学研究所は、2012年3月にスーパーコンピュータシステムを更新しました。新しいスーパーコンピュータシステムはゲノム解析を主な目的とした大規模計算機利活拠点として、最新鋭の大規模クラスター型計算機、大規模メモリ共有型型計算機、および大容量高速ディスク装置で構成されたスーパーコンピュータシステムサービスを提供しています。



- システムハードウェア構成
- システムソフトウェア構成
- システム稼働状況

本サイトは国立遺伝学研究所スーパーコンピュータシステムが提供する計算機リソース、各種アプリケーション、それらの利用方法についての各種情報を提供します。DOBJセンターとして提供する各種サービスについてはDOBJセンターのホームページからご参照ください。

### ディスク利用状況



こちらからどうぞ！

- Webサービス
  - MGAP利用申請
  - MGAPパスワード変更
  - DBJ Pipeline利用申請

現在のUGLキューの利用状況概要です。データは10分置きに更新されます。  
 下記表は、各計算ノードのスロット、メモリの使用率を表します。  
 ジョブ投入にあたり、現在状況の参考にして下さい。  
 利用状況により、新規ジョブは、待ち合わせ時間が発生する可能性があります。