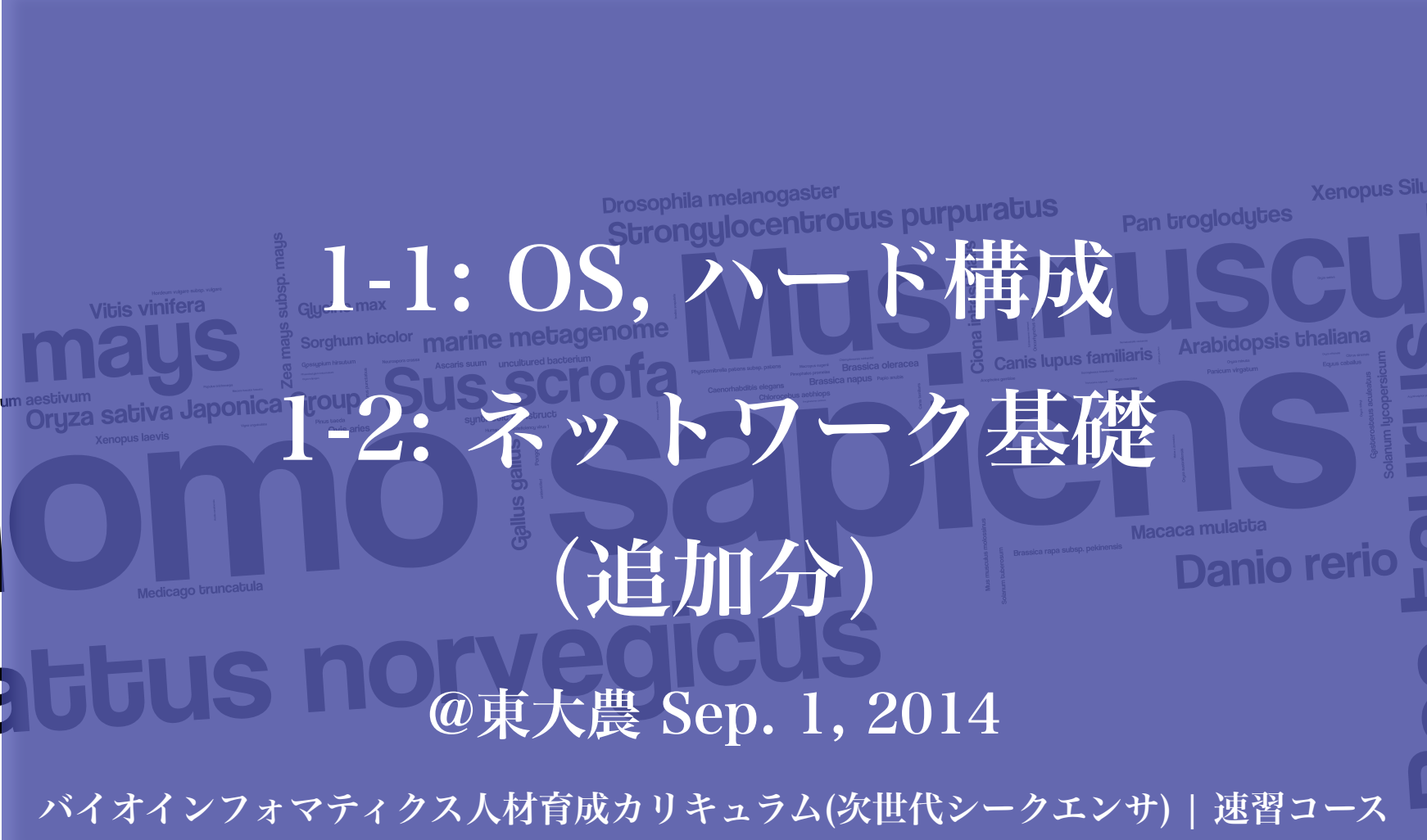


Nicotiana tabacum

Zea mays

Homo sapiens

Rattus norvegicus



1-1: OS, ハード構成

1-2: ネットワーク基礎
(追加分)

@東大農 Sep. 1, 2014

バイオインフォマティクス人材育成カリキュラム(次世代シーケンサ) | 速習コース

中村 保一

NAKAMURA Yasukazu, Professor

国立遺伝学研究所 大量遺伝情報研究室

<http://charles.genes.nig.ac.jp> yn@nig.ac.jp

Bos taurus

- Linuxは、狭義ではLinuxカーネルを意味し、広義ではそれをカーネルとして用いたオペレーティングシステム (OS) を意味する。Linuxカーネルを用いたオペレーティングシステムは、Unix に類似した振る舞いをするので、Unix系のオペレーティングシステムの一つに分類される。

<http://ja.wikipedia.org/wiki/Linux>

UNIX の構成

GNU

Applications

Linux
Mach

Kernel

カーネルは
人間と機械の媒

CPU

Memory

Devices



[リーナス・ベネディクト・トーバルズ \(Linus Benedict Torvalds; 1969-\)](#)

フィンランド、ヘルシンキ出身のプログラマ。Linuxカーネルを開発し、1991年に一般に公開した

- 1991年、ヘルシンキ大学の学生であったリーナス・トーバルズはオペレーティングシステムに好奇心を抱き Intel 80386 CPU を搭載した PC に UNIX 互換の独自のカーネルを開発した。最終的にこれが現在の Linux カーネルへと成長した。

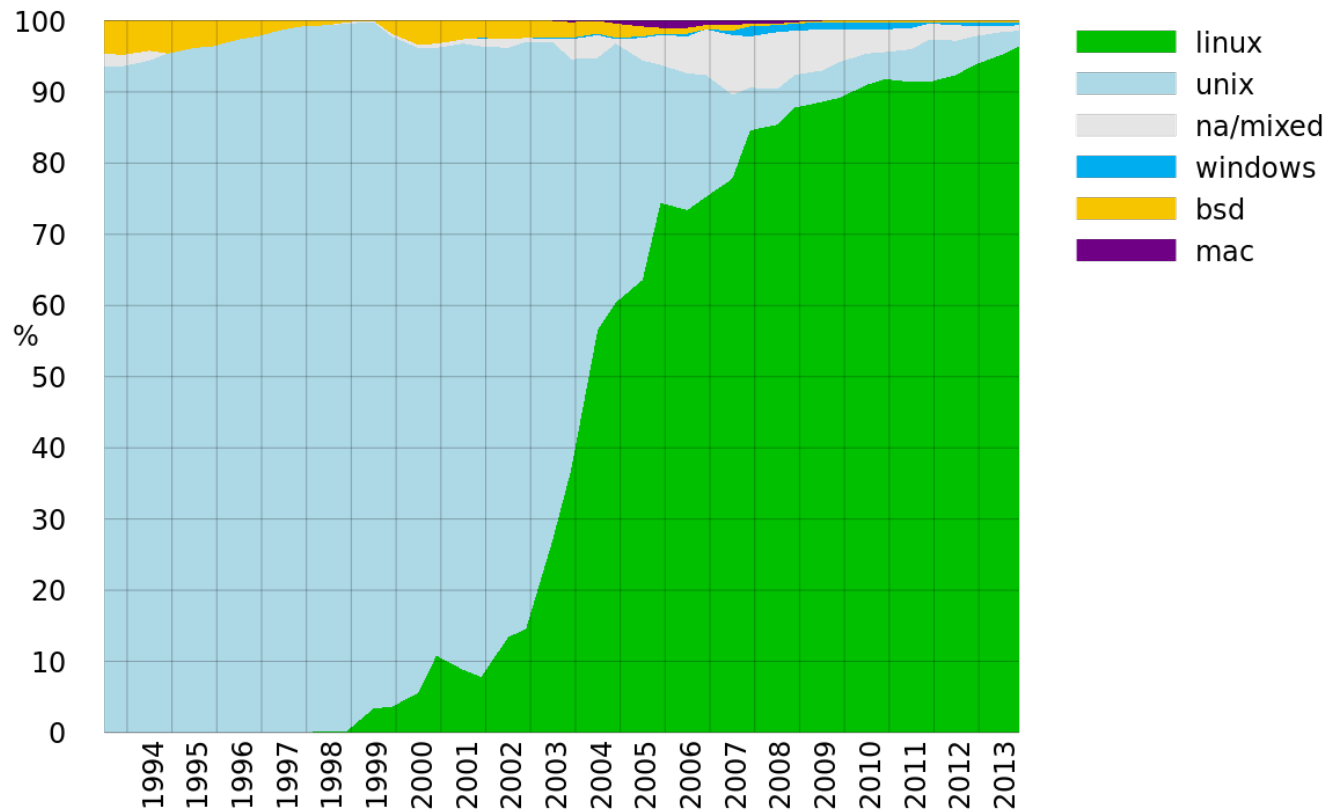
<http://ja.wikipedia.org/wiki/Linux>

見えてるのは実は Linux でなくて GNU

- GNU's Not Unix
- GNU は典型的には Linux と呼ばれるカーネルとともに使われます。この組み合わせが GNU/Linux オペレーティング・システムです。GNU/Linux は何百万人もの人に使われています。間違って“Linux”と呼ばれていますけれども。

<http://www.gnu.org/home.ja.html>

GNU/Linux



[Operating systems used on top 500 supercomputers](#)

解析パイプラインも提供しています

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



DDBJ DNA Data Bank of Japan

Login & Submit | Databases ▾ | English | Contact

Google™ カスタム検索

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

DRA pipeline: ソフトウェア

よく用いられる
解析用ソフトウェアを
用意。クリックだけで
実行可能



ACCOUNT
login ID [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.

Select Query Files
Select Tools
Set QuerySet
Set Genom

Running Status

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"/>	Maq		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	✓		✓				✓					✓	
<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.

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 severe recommend when performing Velvet, total length of those reads is up to 22G bp. Maximum K-mer value is 64.

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



ACCOUNT

login ID [yaskaz]

Logout

Change password

ANALYSIS

Data setup

DRA Start

FTP upload

JOB STATUS

step1.

Preprocessing

step1.

Mapping

step1.

de novo Assembly

step2-All status

HELP

Select Query Files

Select Tools

Set QuerySet

Running Status

Specifying Database of Ref

Major genome sets

Organisms

Genome sets
 TAIR8
 TAIR9
 TAIR10

all.fa

chr01.fa

chr02.fa

chr03.fa

chr04.fa

chr05.fa

chrC.fa

chrM.fa

User original sets

Download or upload reference

Major genome sets

Organisms

Genome sets
 IRGSP Releases Build 4.0
 IRGSP Releases Build 5.0
 IRGSP Releases Build 5.0 masked by RepeatMasker with tigr version5.0
 tigr version6.0
 tigr version6.1
 tigr mitochondrion
 tigr chloroplast

Major genome sets

Organisms

Genome sets
 Homo sapiens Feb. 2009 (hg19)
 Mar.2006 (hg18)
 May.2004 (hg17)
 NCBI build 36.1_CRA
 NCBI build 36.1_Celera
 NCBI build 36.1_ref
 NCBI build 36.2_CRA
 NCBI build 36.2_Celera
 NCBI build 36.2_ref
 NCBI build 36.3_CRA
 NCBI build 36.3_Celera
 NCBI build 36.3_ref
 NCBI build 36.3_HuRef
 NCBI build 37.1_CRA
 NCBI build 37.1_Celera
 NCBI build 37.1_GRCh
 NCBI build 37.1_HuRef

DDBJ パイプラインの解析の流れ

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



DDBJ Read Annotation Pipeline

English

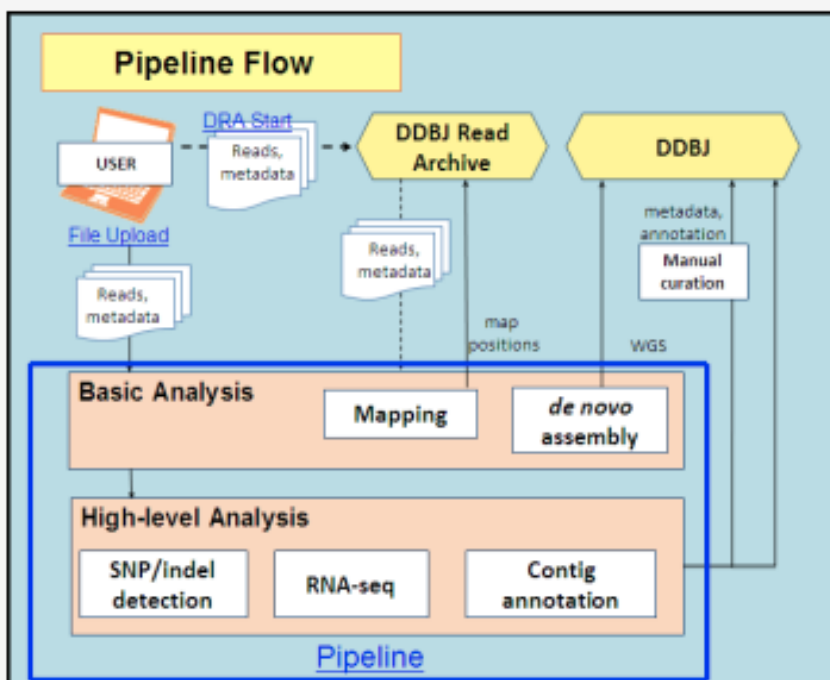
Japanese

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

LOGIN

New account

Login as "guest"



User ID:

ゲストとして
ログイン

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

アップロードされている配列

Account: login ID [guest], Logout

Analysis: Data setup, DRA Start, FTP upload, HTTP upload, DRA Import, Preprocessing Start

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

Navigation: Select Query Files, Select Tools, Options, Confirmation

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.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
<input type="checkbox"/>	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

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 GenomeSet → Set Map Options → Confirmation →

Running Status

Select

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"/>	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

DELETED NEXT

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 → 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

[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

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

Selecting Query Files

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

Metadata of the DRA entry.

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

Select a metadata

STUDY TITLE	Whole genome sequencing of <i>Baillus subtilis</i> subsp. <i>natto</i> 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 checked="" 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.)

DELETED 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 /
de novo Assembly

step-2

WorkflowGenome (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"/>	Maq		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	✓		✓				✓					✓	
<input type="checkbox"/>	Bowtie2		2.0.0	✓	✓	✓				✓	✓				✓	For reads longer than 100bp, use the -x option.

 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		1.2.03			✓	✓	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		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
<input checked="" type="radio"/>	BLAT	Single-end analysis only

BACK

NEXT

velvet で
アセンブル
しましょう

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

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 velveth.
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

The screenshot shows the 'Run Confirmation' page on the DDBJ website. The page has a navigation bar at the top with steps: 'Select Query Files', 'Select Tools', 'Set QuerySet', and 'Run Confirmation'. On the left, there is a sidebar with sections: '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), and 'HELP' (HELP, TUTORIAL, Contact Us). The main content area is titled 'Run Confirmation' and includes a 'Destination of mail' field with the email 'yn@nig.ac.jp' circled in red. Below this is a table for 'Query sets' with columns: PairedOrientation, RunAccession, RunAlias, RowLength, QualityScore1, QualityScore2. The table contains one row: paired, DRR000001, DRR000001, 36. Below the table are sections for 'Assembly commands' and 'Set optional parameters of the paired-end analysis'. The 'Set optional parameters' section includes: 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. At the bottom, there is a 'Set filtered length for contigs' section with a checkbox and a text input field.

連絡先いれたら
実行可能

でも今は
押さないで！

「RUN を押した」と思ってください



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

Preprocessing (unt)

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

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"/>	4914	guest	DRA000001 DRR000001	P	complete	SOAPdenovo	9,977,388	36	<input type="button" value="View"/>		2013-01-11 22:06:40	2013-01-11 22:13:40	00:06:59
<input type="checkbox"/>	4912	guest	--- test.txt	S	complete	SOAPdenovo	1	---	<input type="button" value="View"/>		2013-01-11 18:01:34	2013-01-11 18:02:16	00:00:42
<input type="checkbox"/>	4911	guest	--- preprocessing	S	error	SOAPdenovo		---	<input type="button" value="View"/>		---	---	
<input type="checkbox"/>	4909	---	--- HPS1	P	complete	ABYSS	5,754,246	---			2013-01-11 12:14:34	2013-01-11 13:52:21	01:37:47
<input type="checkbox"/>	4908	guest	DRA000001 2008-09-12.BES	P	complete	Velvet	9,977,388	36	<input type="button" value="View"/>		2013-01-11 12:13:53	2013-01-11 16:10:01	03:56:08
<input type="checkbox"/>	4907	---	--- HPS1	P	complete	ABYSS	5,754,246	---			2013-01-11 12:03:54	2013-01-11 13:24:40	01:20:45
<input type="checkbox"/>	4900	---	--- lon_mt20mergac	S	complete	ABYSS	148,666	---			2013-01-10 15:32:13		00:16:28

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.
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

Mappingの例 (DRAsearch+pipeline)

シロイヌナズナ
alternative splicing

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

Keyword :

Show records Sort by

Search Results (2263 records)

/ 114

Filtered by

document type:study(1161) sample(631) experiment(369) submission(75) run(26) analysis(1)
organism:Arabidopsis thaliana(1000) Mus musculus(339) Homo sapiens(335) Drosophila melanogaster(40)
Saccharomyces cerevisiae(22) Arabidopsis lyrata(19)

#	META_FILE	ACCESSION	STUDY	STUDY_TITLE	STUDY_TYPE	ORGANISM	BASES	SUBMITTED	CENTER_NAME
1	SRA050132.study.xml <?xml version="1.0" encoding="UTF-8"?> <STUDY center_name="NCSU" alias="Arabidopsis - Pseudomonas alternative splicing study" accession="SRP010938"> </STUDY>	SRP010938	SRP010938	Arabidopsis thaliana strain:Columbia (Col-0) Transcriptome or Gene expression	Transcriptome Analysis	Arabidopsis thaliana	85.2G	2012-02-15	NCSU
2	SRA009031.study.xml <?xml version="1.0" encoding="UTF-8"?> <SUBMITTER_ID>SRA009031</SUBMITTER_ID> <IDENTIFIERS><DESCRIPTOR><STUDY_TITLE>Transcriptome-wide map of alternative splicing in Arabidopsis</STUDY_TITLE></IDENTIFIERS></SUBMITTER_ID>	SRP000935	SRP000935	Transcriptome-wide map of alternative splicing in Arabidopsis	Transcriptome Analysis	Arabidopsis thaliana	12.5G	2009-07-07	OSU-CGRB
3	SRA050132.submission.xml <?xml version="1.0" encoding="UTF-8"?> <SUBMISSION accession="SRP010938" alias="Arabidopsis alternative splicing project"></SUBMISSION>	SRA050132	SRP010938	Arabidopsis thaliana strain:Columbia (Col-0) Transcriptome or Gene expression	Transcriptome Analysis	Arabidopsis thaliana	85.2G	2012-02-15	NCSU
4	SRA044892.study.xml <?xml version="1.0" encoding="UTF-8"?> <STUDY center_name="Institute of Plant and Microbial Biology, Academia Sinica" alias="Genome-wide detection of context-sensitive alternative splicing in Arabidopsis roots"></STUDY>	SRP007763	SRP007763	Genome-wide detection of context-sensitive alternative splicing in Arabidopsis roots	Other	Arabidopsis thaliana	1.3G	2011-08-11	Institute of Plant and Microbial Biology, Academia

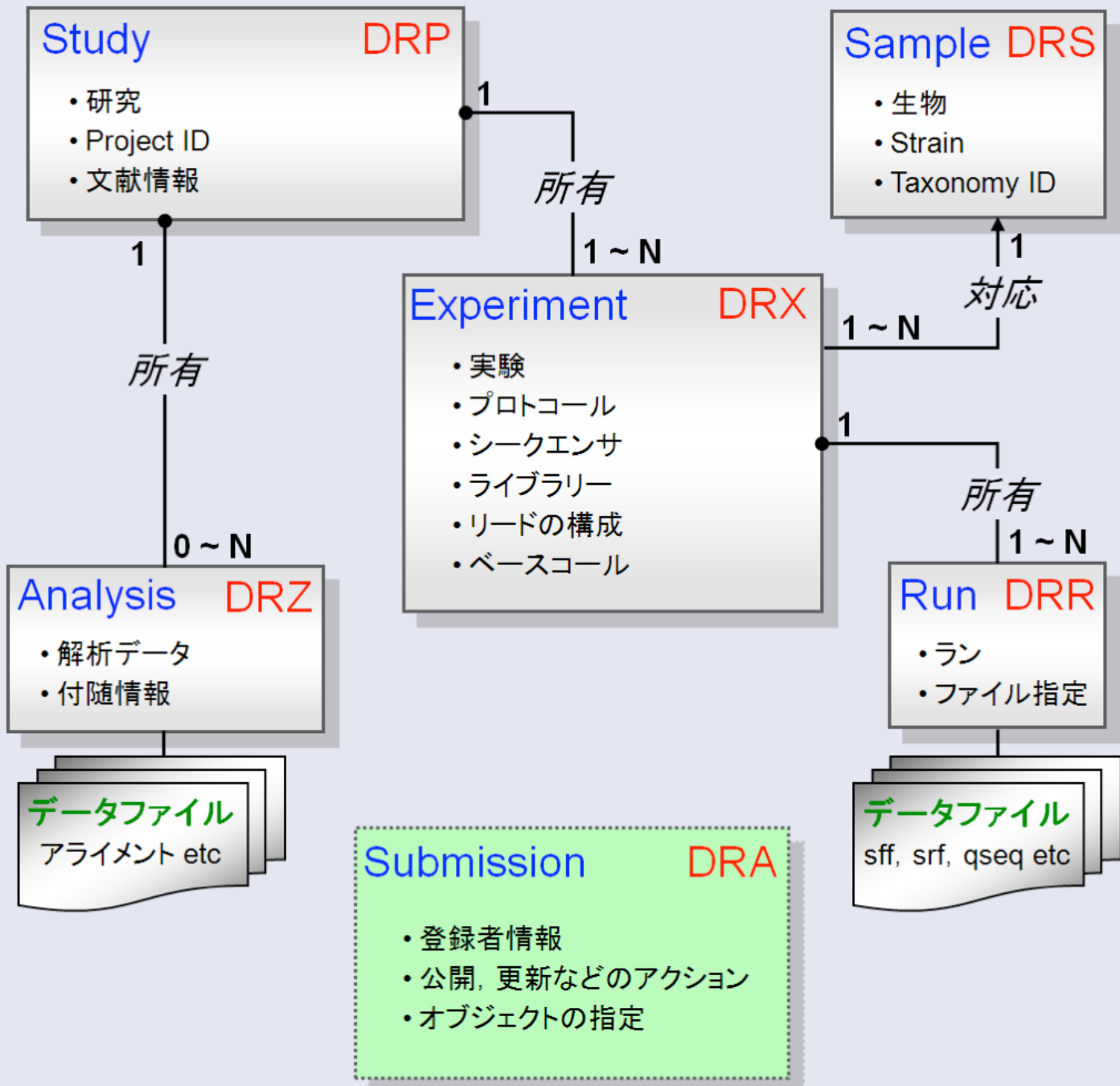
データのIDはこちら

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 .. [more]
Description	
Center Name	Institute of Plant and Microbial Biology, Academia

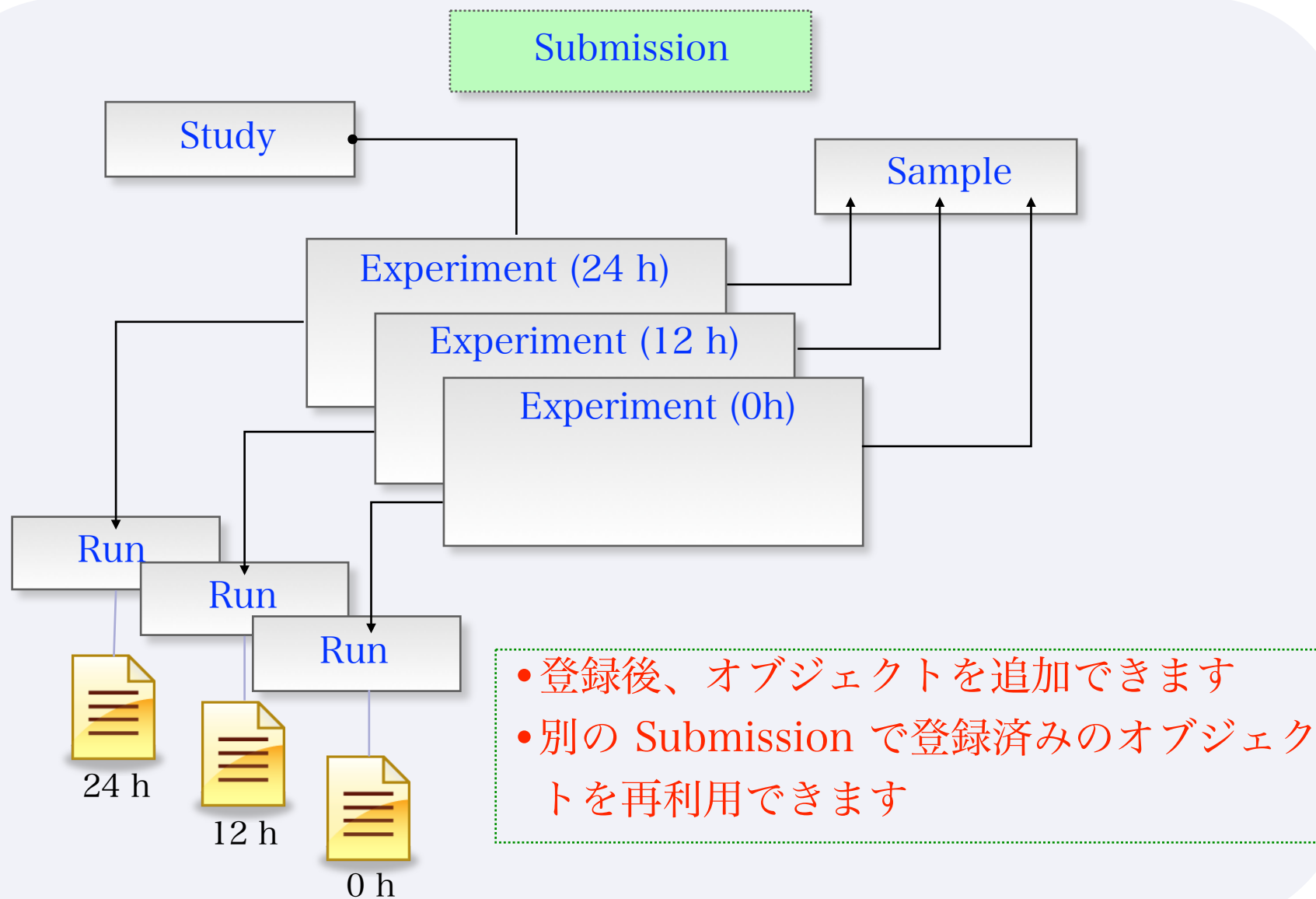
Navigation	
Submission	SRA044892 FTP
Experiment	SRX092040 FASTQ SRA
Sample	SRS256250

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



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

step-1
Preprocessing
Mapping / de novo Assembly

step-2
Workflow
Genome (SNP/Short Indel)
RNA-seq (Tag o
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](#)

(in not select.)

DRA entry tab.
entry is added in "Private DRA entry" tabs.
please retry it.

done : file is ready, failed : please retry it, preparing : file is not yet in

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	DRA000593	2012-08-20 14:04:46.09

でも今は押さないで！

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 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.)

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

WorkflowGenome (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

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	<input type="button" value="Download"/>	<input type="button" value="View"/>
Sample	SRS256250 SRS256251 SRS256252	Control Iron Phosphate	SRA044892.sample.xml	<input type="button" value="Download"/>	<input type="button" value="View"/>
Study	SRP007763	Alternative Splicing	SRA044892.study.xml	<input type="button" value="Download"/>	<input type="button" value="View"/>
Experiment	SRX092046	control	SRA044892.experiment.xml	<input type="button" value="Download"/>	<input type="button" value="View"/>
Run	SRR331219 SRR331224	control iron	SRA044892.run.xml	<input type="button" value="Download"/>	<input type="button" value="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.

	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



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

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"/>	Maq		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	✓		✓				✓				✓	

<input checked="" 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.
-------------------------------------	-------------------------	--	-------	---	---	---	--	--	--	---	---	--	--	---	--

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 severe 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'

<input type="checkbox"/>	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:
all check
✓ TAIR8
TAIR9
TAIR10

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



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 --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 -d1000000 -f refgenome.fasta out.bam | bcftools view -bvcg -> out.var.raw.bcf
- bcftools view out.var.raw.bcf | vcftutils.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

The screenshot shows the DDBJ Run Confirmation page. The page title is "Run Confirmation". The "Destination of mail" field contains "yn@nig.ac.jp", which is circled in red. A speech bubble above the "RUN" button says "連絡先いれたら 実行可能" (If you enter the contact information, execution is possible). Another speech bubble below it says "でも今は 押さないで!" (But don't press now!). The "RUN" button is also circled in red. The page includes a navigation bar with "Select Query Files", "Select Tools", "Set QuerySet", and "Run". The left sidebar has sections for "ACCOUNT", "ANALYSIS", "JOB STATUS", and "HELP". The main content area shows "Query sets" with a table, "genome sets" (TAIR10), and "Command Options" for bowtie2.

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 → Run

Running Status

Run Confirmation

Destination of mail

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

yn@nig.ac.jp

Reference Genome Map [bowtie2]

Query sets

Query set1

PairedOrientation	RunAccession	RunAlias	RowLength	QualityScore1	QualityScore2
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

BACK RUN

連絡先いれたら 実行可能

でも今は 押さないで!

「RUN を押した」と思ってください

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
step2-All status

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

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"/>	4915	guest	SRA044892 control	S	running	Bowtie2	5,925,048	---	121 M	<input type="button" value="View"/>	2013-01-11 23:23:03	
<input type="checkbox"/>	4913	guest	SRA049447 9870	S	complete	Bowtie2	14,278,727	---	4 M	<input type="button" value="View"/>	2013-01-11 18:21:55 2013-01-11 19:26:40	01:04:44
<input type="checkbox"/>	4910	---	ERA000092 1_dpfi_assay1_1 1_dpfi_assay1_2 1_dpfi_assay1_3 1_dpfi_assay2_1 and more...	P	running	Bowtie2	375,421,197	---	1,379 M		2013-01-11 16:19:14	
<input type="checkbox"/>	4906	---	A1_Unshu_Paire	P	running	bwa	110,759,316	---	299 M		2013-01-10 16:24:05	
<input type="checkbox"/>	4905	---	lon_mt20mergac	S	complete	Bowtie2	148,666	---	16,520		2013-01-10 15:58:39 2013-01-10 16:04:15	00:05:36
<input type="checkbox"/>	4904	---	lon_mt20mergac	S	complete	Bowtie2	148,666	---	16,831		2013-01-10 15:54:22 2013-01-10 16:00:46	00:06:24
<input type="checkbox"/>	4903	---	---	S	complete	Bowtie2	148,666	---	16,589		2013-01-10	00:05:36

処理状況は
こちらから

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	Read length	Alias
SRR331219	SRR331219.fastq.bz2	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
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

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

実行結果

Development Team.