

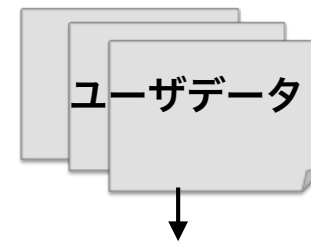
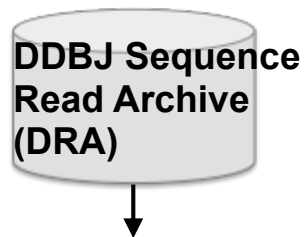
2017年度 「先進ゲノム支援」 情報解析講習会

# DDBJスパコンでの解析の実践I (DDBJパイプライン)

国立遺伝学研究所 大量遺伝情報研究室

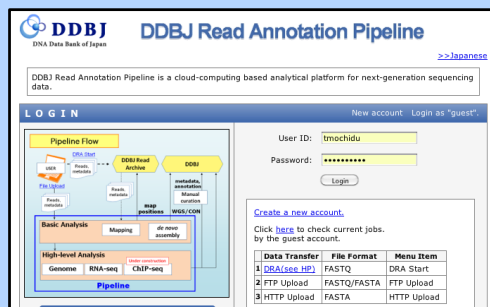
望月孝子

# DDBJ Read Annotation Pipeline 全体像



## DDBJ Pipeline: DDBJ Read Annotation Pipeline

### 基礎処理部

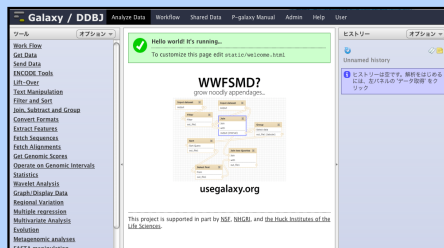


マッピング

de novo アセンブリ

### 高次処理部

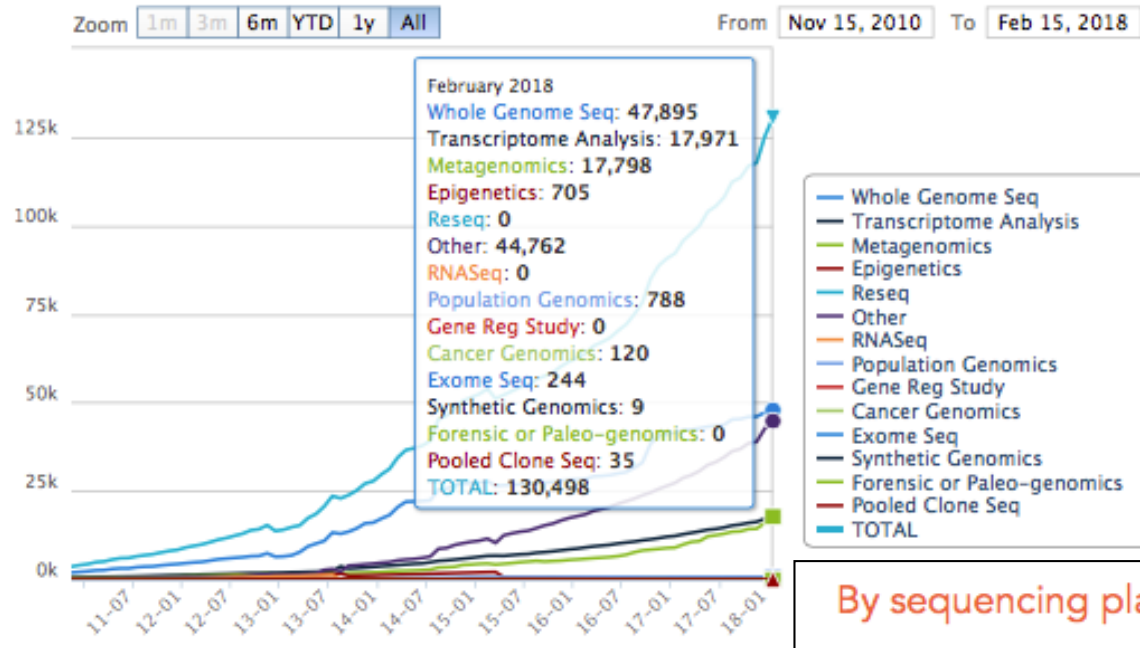
### 解析目的別ワークフロー



DNA多型  
注釈  
(DNApod)

# SRA データ登録状況

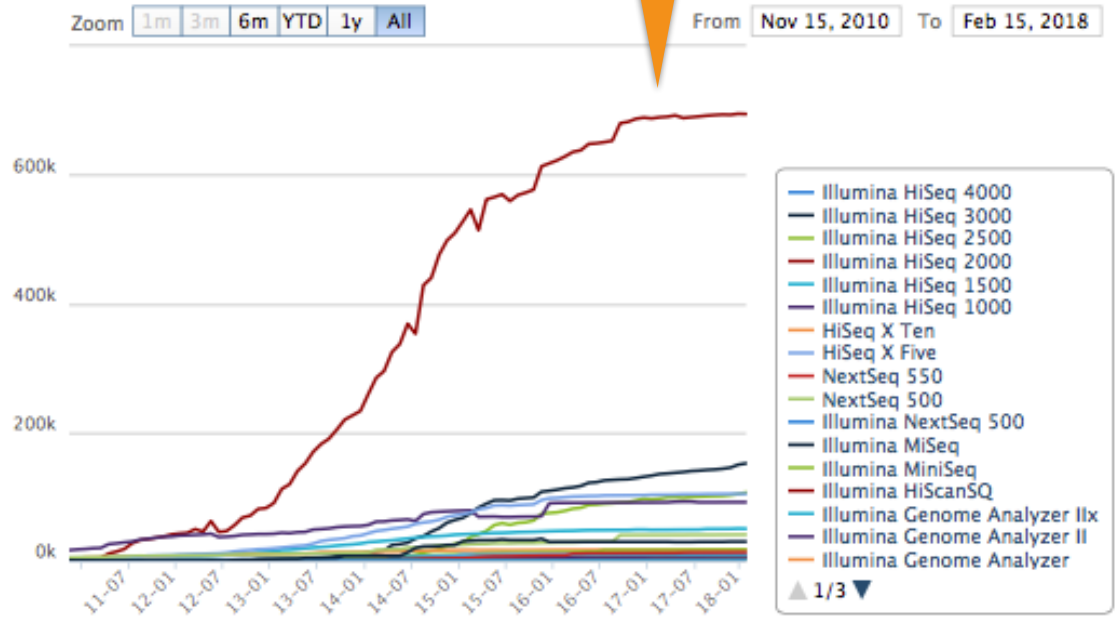
## By study types



2018/02 時点

Illumina HiSeq 2000: 694,507

## By sequencing platforms



DBCLS SRA

DBCLS SRA

<http://sra.dbcls.jp/trends.html>

# SRAデータサーチ

DBCLS SRA metadata search (beta) Home Blog

## DBCLS SRA Metadata Search project: Soylatte

Setting search filter

Sample Organism

Study Type

Sequencing Instrument

Fulltext search

search without filtering

<http://sra.dbcls.jp/search>

POWERED BY DBCLS, PROJECT LICENSED UNDER CC-BY 2.1 JP / FONT AWESOME BY DAVE GANDY - [HTTP://FONTAWESOME.GITHUB.COM/FONT-AWESOME](http://fontawesome.github.com/font-awesome)

EMBL-EBI Services

## ENA European Nucleotide Archive

Examples: BN000065, histone

Home Search & Browse Submit & Update Software About ENA Support

Advanced  Upload accession

Search query

Select content:

- Assembly
- Sequence
- Contig set
- Coding
- Non-coding
- Read
- Analysis
- Study
- Sample

<https://www.ebi.ac.uk/ena/data/warehouse/search>

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

## DRASearch

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

Accession :

Organism :  StudyType :

CenterName :  Platform :

Keyword :

Show  records Sort by

Data Last Update 2018-03-05

# DDBJ pipeline 提供ツール一覧

## 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>		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 type="checkbox"/>	<a href="#">Bowtie2</a>		2.2.6	✓	✓	✓	✓	✓	✓	✓				✓	For reads longer than about 50 bp, Bowtie2 is generally faster, more sensitive, and uses less memory than Bowtie1.
<input type="checkbox"/>	<a href="#">TopHat2</a>		2.1.0	✓		✓	✓	✓	✓	✓				✓	

## de novo Assembly

	Tool	Help	Version	Base space	Color space	Paired-end	MSS(WGS)	Comment
<input type="checkbox"/>	<a href="#">SOAPdenovo</a>		2.04-r240	✓		✓		
<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>		2.1.1	✓		✓		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)
<input type="checkbox"/>	<a href="#">Canu</a>		1.6					Nanopore only (Beta version)

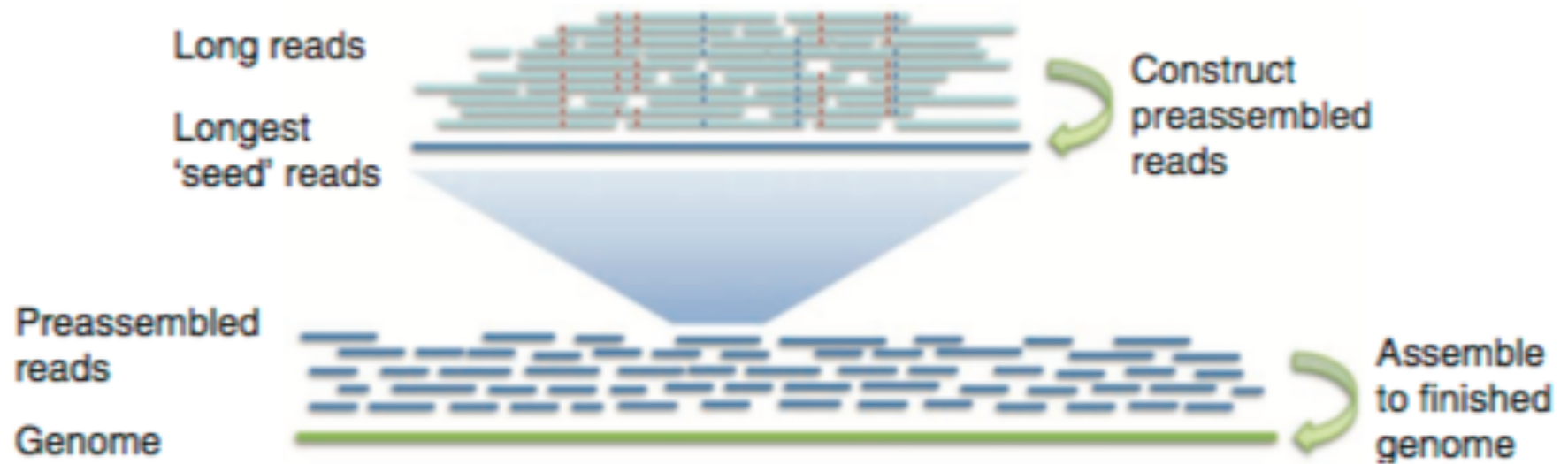
今回の講習はHGAP を使用します。

## 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

# HGAP の処理の流れ



Chin C. S. *et al.*, Nature Methods, 10, 563–569, 2013

## (1) Preassembly

一定以上の長さのリード (シード) に、短いリードをマップし、エラー補正を行う。

## (2) Assembly

Preassemble reads をアセンブルする。

## (3) Consensus Polishing

コンティグの再補正を行う。

# HGAP version 2.2.0 実行方法

---

## 入力ファイル

bax.h5 形式ファイル (DDBJ SRA では公開されていない)

1セルにつき 3 つの bax.h5 ファイル

## パラメータ

GenomeSize: 推定ゲノムサイズ

Minimum Seed length: デフォルト 6000 及び Automatic Estimation

カバレッジが  $\times 25$  以上あるのであれば、 $25x$  のリードでシードリードが補正されるように自動で最小のシードリード長を計算してくれる。もし、カバレッジが  $\times 25$  よりも小さければ指定した値を最小のシードリード長にする。

## 出力ファイル

アセンブル結果ファイル (fasta / fastq)

補正されたロングリードファイル

ログファイル

## 講習用データ

乳酸菌 *Lactobacillus hokkaidonensis* LOOC260<sup>T</sup>

PacBio RS II で解読

# アクセス

## キーワード検索



The image shows a Google search interface. The search bar contains the text "DDBJ pipeline". Below the search bar, there are navigation tabs: "すべて" (All), "ニュース" (News), "動画" (Videos), "画像" (Images), "地図" (Maps), "もっと見る" (More), "設定" (Settings), and "ツール" (Tools). The search results show approximately 54,300 items found in 0.45 seconds. The top result is "DDBJ Read Annotation Pipeline" with the URL <https://p.ddbj.nig.ac.jp/>. Below the title, there is a snippet: "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。". Below this, there is another result titled "DDBJ Read Annotation Pipeline の紹介と実習@第30回 DDBJing 講習 ..." with the URL [www.ddbj.nig.ac.jp/movie/44643.html](http://www.ddbj.nig.ac.jp/movie/44643.html). The snippet for this result reads: "DDBJ Read Annotation Pipeline の紹介と実習@第30回 DDBJing 講習会 in 東京. 2014年12月18日(木)に行われた第30回 DDBJing 講習会 in 東京 での、長崎 英樹 研究員 (国立遺伝学研究所 大量遺伝情報研究室) による「DDBJ Read Annotation Pipeline の紹介と実習」についての講義です。 (時間: 約1時間) . 2015-01-04 ...". At the bottom, there is a link for "Pipeline ヘルプ - DDBJ".

Google

DDBJ pipeline

すべて ニュース 動画 画像 地図 もっと見る 設定 ツール

約 54,300 件 (0.45 秒)

**DDBJ Read Annotation Pipeline**  
<https://p.ddbj.nig.ac.jp/> ▼ このページを訳す  
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。

**DDBJ Read Annotation Pipeline の紹介と実習@第30回 DDBJing 講習 ...**  
[www.ddbj.nig.ac.jp/movie/44643.html](http://www.ddbj.nig.ac.jp/movie/44643.html) ▼  
DDBJ Read Annotation Pipeline の紹介と実習@第30回 DDBJing 講習会 in 東京. 2014年12月18日(木)に行われた第30回 DDBJing 講習会 in 東京 での、長崎 英樹 研究員 (国立遺伝学研究所 大量遺伝情報研究室) による「DDBJ Read Annotation Pipeline の紹介と実習」についての講義です。 (時間: 約1時間) . 2015-01-04 ...

[Pipeline ヘルプ - DDBJ](#)



# ログイン

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

English Japanese

LOG IN New account Login as guest

1. ユーザIDとパスワードを設定

User ID: koshu40

Password: \*\*\*\*\*

Login

Check current jobs  
\* by the guest account.

Manual & tutorial

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

各自の講習 ID でログインしてください。

アカウントの作成はこちらから。

## Registration form for pipeline user accounts

Note that this account is NOT registered as a NIG supercomputer account.

As DDBJ Pipeline is a webservice of NIG supercomputer, user information was publicly opened to the internet from here. ([Supercomputer User Policy](#))

After registration, you will receive a confirmation email with your user ID and initial password. Please input your email address correctly.

* UserID:	<input type="text"/>	Use 6 to 16 characters.
* Email address:	<input type="text"/>	
* Retype email address:	<input type="text"/>	* for confirmation.
* First name:	<input type="text"/>	
* Last name:	<input type="text"/>	
* Institution with department:	<input type="text"/>	ex. Center for Information Biology, National Institute of Genetics.
* Country:	<input type="text" value="AFGHANISTAN"/>	
* Address:	<input type="text"/>	ex. 1111 Yata, Mishima, Shizuoka
* Postal/Zip code:	<input type="text"/>	ex. 411-8540
* Telephone number:	<input type="text"/>	ex. +81559816859
* Purpose of utilization:	<input type="text"/>	

\* All contents are required.

Registration

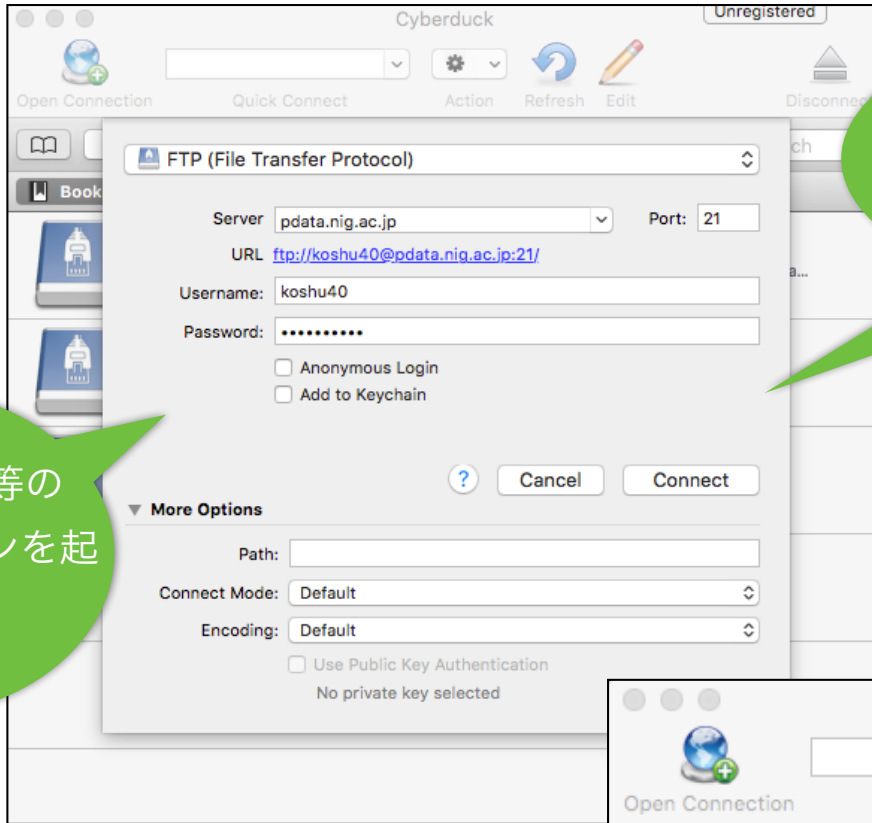
[<< Back to login page](#)

2. Login をクリック



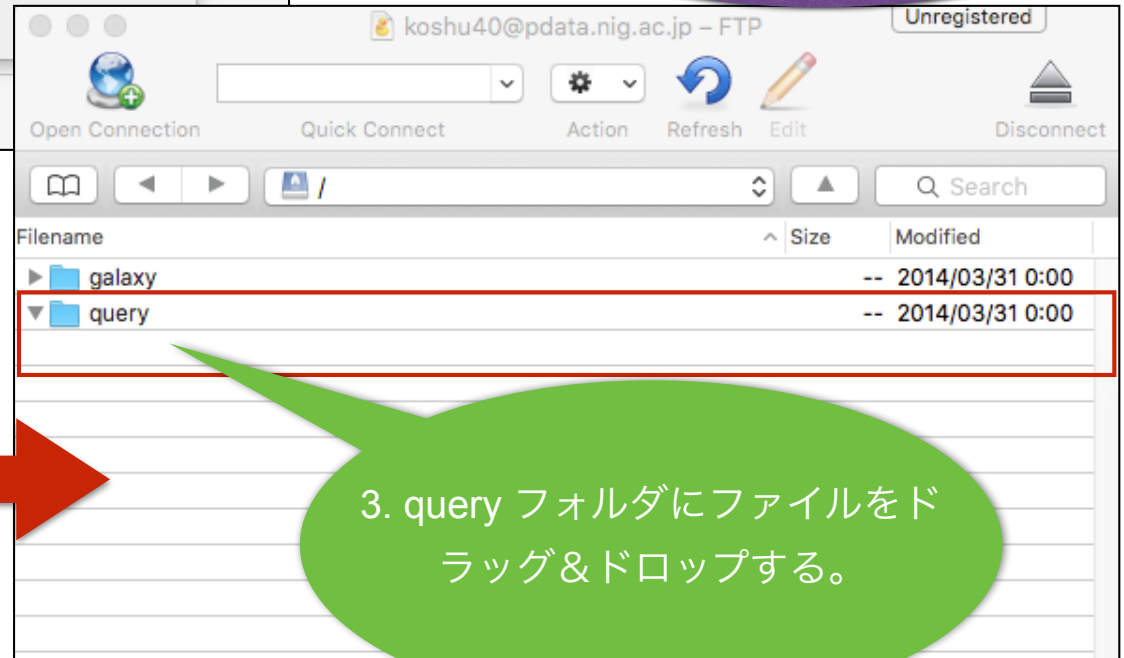
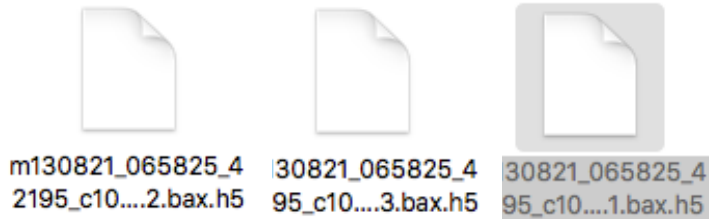
Cyberduck

1. Cyberduck 等のアプリケーションを起動



2. FTP の情報を設定する。  
サーバー名、ポート番号、ユーザID、パスワード

講習では 1 セル分のデータを使用。  
このステップは準備済みです。



3. query フォルダにファイルをドラッグ&ドロップする。

**DDBJ**  
DNA Data Bank of Japan

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

Running Status

**ACCOUNT**  
login ID [koshu40]  
Logout  
Change pass

**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

**Query Files**  
FASTQ file. ERR005143(forward)  
FASTQ file. ERR005143(reverse)

1. FTP upload を  
クリック

2. Add new files を選択

FTP upload Private DRA entry Import public DRA Preprocessing HTTP upload

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

Select All Clear All

Filename	Description	Layout	Instrument model	File size
----------	-------------	--------	------------------	-----------

DELETE NEXT

## Registration of fastq/fasta files

1. Upload FASTA/FASTQ files 2. Select FASTA/FASTQ files 3. Registration

**Please upload query files.**

To use your fasta or fastq files as pipeline query, you need to upload files to our server via FTP or HTTP. FTP uploading works faster than HTTP uploading. Therefore we recommend using FTP rather than HTTP

**By FTP (Recommended)**

**FTP Configuration.**

Server : Port	pdata.nig.ac.jp:21
Security	SSL Explicit encryption
User ID/password	Your Pipeline login ID/password If you can't login via FTP, retry after <a href="#">changing password</a> .

[FTP setting manual \(English\)](#) [FTP setting manual \(Japanese\)](#)

**Recommended FTP client softwares.**

Windows	<a href="#">FFFTP</a> <a href="#">WinSCP</a>
Mac OS X	<a href="#">Cyberduck</a>
LinuxOS	<a href="#">FileZilla</a>

For security our FTP server utilizes FTP over SSL protocol (FTPS). Other FTP client softwares can be used if they support FTPS.

**NOTICE**

- Uploaded files **cannot be seen** from other Pipeline users.
- when you connected to FTP server, there are two directories, "query" and "galaxy". **Please upload into "query" directory.** If you uploaded to same level as the "query" directory, the file cannot be used in Pipeline.
- The uploaded files will be displayed in the list below after a few minutes. (It takes 2-5 min per 1GB) When uploading is completed, files are transferred to Pipeline data directory from FTP server. **So files seem to be removed, but it is normal operation.**
- Please ensure that uploading files have appropriate file extensions.  
eg. In the case of Bzip2 files, please add the ".bz2" extension.

**Supported file type**

Filetype	Extension
Plain text	.fasta, .fq .fastq, .fa etc...
Gzip	.gz
Bzip2	.bz2

Bzip2 is recommended, because save disk space usage and transfer.

**By HTTP (slower)**

If you can't use FTP uploading, click "Browse and Upload" button and select FASTA/FASTQ files to be uploaded.

[Browse and Upload](#) [Delete Files](#)

	filename	type	size	timestamp
<input type="checkbox"/>	m130821_065825_42195_c100539522550000001823089611241356_s1_p0.1.bax.h5	invalid	784.3 MB	2018-03-05 01:18:32
<input type="checkbox"/>	m130821_065825_42195_c100539522550000001823089611241356_s1_p0.2.bax.h5	invalid	803.9 MB	2018-03-08 03:30:56
<input type="checkbox"/>	m130821_065825_42195_c100539522550000001823089611241356_s1_p0.3.bax.h5	invalid	945.5 MB	2018-03-08 03:29:22

Go to the next page after uploading files.

[Next STEP >](#)

講習用のデータが3ファイルあるのを確認してください。

1. Next STEP を  
クリック

# FTP ファイルアップロード

## Step 4

**Registration of fastq/fasta files**

1. Upload FASTA/FASTQ files 2. Select FASTA/FASTQ files 3. Registration

Please specify read layout to uploaded files.

1. Select a read layout:

Read layout :

2. Select a FASTA/FASTQ file:

If you are select Paired-end, please specify

filename	type	size	timestamp
<input type="radio"/> Not select			
<input checked="" type="radio"/> m130821_065825_42195_c100539522550000001823089611241356_s1_p0.1.bax.h5	invalid	784.3 MB	2018-03-05 01:18:32
<input type="radio"/> m130821_065825_42195_c100539522550000001823089611241356_s1_p0.2.bax.h5	invalid		2018-03-08 00:56
<input type="radio"/> m130821_065825_42195_c100539522550000001823089611241356_s1			

1. Select a read layout : Single-end を選択

2. ファイルを選択

3. Next STEPをクリック

Are you sure you want to submit ?

6. OK をクリック

キャンセル

OK

**Registration complete.**

Press "Mapping / Assembly" button to go to job input pages.

Assembly / Mapping

**Registration of fastq/fasta files**

1. Upload FASTA/FASTQ files 2. Select FASTA/FASTQ files 3. Registration

Please specify instrument model.

SelectedFile 1 | m130821\_065825\_42195\_c100539522550000001823089611241356\_s1\_p0.1.bax.h5

SelectedFile 2 | Not select

Read layout | Single-end

Instrument model |

(Required) Study title |

NOTICE: After confirming your entries, push the SUBMIT button to register uploaded files.

4. Instrument model: PacBio を選択

5. Study title を入力

6. SUBMIT をクリック

7. Step 2 - 4 を3ファイル分行う。

# クエリの選択

**DDBJ**  
DNA Data Bank of Japan

**ACCOUNT**  
login ID [koshu40]  
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

1. FTP upload をク  
リック

## Selecting Query Files

- Missing FASTQ file. ERR005143(forward)
- Missing FASTQ file. ERR005143(reverse)

NEXT

FTP upload Private DRA entry Import public DRA Preprocessing HTTP upload

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

Select All Clear All

Filename	Description	Layout	Instrument model	File size
<input checked="" type="checkbox"/> m130821_065825_42195_c100539522550000001823089611241356_s1_p0.3.bax.h5	Lactobacillus hokkaidonensis LOOC260T 3	single	PacBio	945.5 MB
<input checked="" type="checkbox"/> m130821_065825_42195_c100539522550000001823089611241356_s1_p0.2.bax.h5	Lactobacillus hokkaidonensis LOOC260T 2	single	PacBio	803.9 MB
<input checked="" type="checkbox"/> m130821_065825_42195_c100539522550000001823089611241356_s1_p0.1.bax.h5	Lactobacillus hokkaidonensis	single	PacBio	784.3 MB

2. ファイルを選択

2.NEXTをクリック

DELETE NEXT

# ツールの選択

## 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>		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 type="checkbox"/>	<a href="#">Bowtie2</a>		2.2.6	✓	✓	✓	✓	✓	✓	✓				✓	For reads longer than about 50 bp, Bowtie2 is generally faster, more sensitive, and uses less memory than Bowtie1.
<input type="checkbox"/>	<a href="#">TopHat2</a>		2.1.0	✓		✓	✓	✓	✓					✓	

1. *de novo* Assembly を選択

### *de novo* Assembly

	Tool	Help	Version	Base space	Color space	Paired-end	MSS(WGS)	Comment
<input type="checkbox"/>	<a href="#">SOAPdenovo</a>		2.04-r240	✓		✓		
<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>		2.1.1	✓		✓		RNA-Seq De novo Assembly
<input type="checkbox"/>	<a href="#">Platanus</a>		1.2.2	✓		✓		
<input checked="" 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)
<input type="checkbox"/>	<a href="#">Canu</a>		1.6					Nanopore only (Beta version)

2. HGAP を選択

# クエリセットの作成

## Generating Query Sets from Query Read Files

RESET BACK NEXT

Single analysis  
Layout of single sequence.

5' 3'

Linker(1) Target Linker(2)

1. ファイルを選択

	Run ACCESSION	Read length	Quality Score
<input checked="" type="checkbox"/>	m130821_065825_42195_c100539522550000001823089611241356_s1_p0.3.bax.h5	bp	
<input checked="" type="checkbox"/>	m130821_065825_42195_c100539522550000001823089611241356_s1_p0.2.bax.h5	bp	
<input checked="" type="checkbox"/>	m130821_065825_42195_c100539522550000001823089611241356_s1_p0.1.bax.h5	bp	

confirm

2. confirmをクリック

QUERY SET

RESET BACK NEXT

## Generating Query Sets from Query Read Files

RESET BACK NEXT

Single analysis  
Layout of single sequence.

5' 3'

Linker(1) Target Linker(2)

Run ACCESSION Read length Quality Score

3. NEXTをクリック

QUERY SET  
Query set1

PairedOrientation	RunAccession	RunAlias	RowLength	QualityScore1	QualityScore2
single	30838	Lactobacillus hokkaidonensis LOOC260T 3			
single	30837	Lactobacillus hokkaidonensis LOOC260T 2			
single	30836	Lactobacillus hokkaidonensis LOOC260T 1			

RESET BACK NEXT



# パラメータの設定

## Setting for De Novo Assembly

BACK NEXT

### hgap

#### Set optional parameters for HGAP pipeline

Select UGE-node to run :

month\_fat (32 CPUs and 320GB memory)

month\_medium (32 CPUs and 256GB memory)

Same results will be generated with either option.  
You can check the CPU and memory usage at [NIG-SC Website](#).

1 : The approximate genome size, in base pairs.(Must be a value between 1 and 150000000)

GenomeSize =

2 : The minimum length of reads (in base pairs) to use as seeds for pre-assembly.

Minimum Seed Length :

Automatic Estimation

If the coverage exceeds 25X, the Minimum Seed Read Length that results in at least 25X coverage by the longest subreads will be calculated automatically. If the coverage is less than 25X, the user-specified value will be used.

Use Manually Specified Value (regardless of the coverage)

BACK NEXT

1.パラメータの値  
を設定

講習では乳酸菌の一般的な  
ゲノムサイズを指定

2.NEXTの値を設定

# 実行の確認

**Run Confirmation**

Destination of mail

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

xxxx@nig.ac.jp \* Required

Result files will be deleted 60 days after submission.

Assembly [hgap]

Query sets

Query set1

PairedOrientation	RunAccession	RunAlias	RowLength	QualityScore1	QualityScore2
single	30838	Lactobacillus hokkaidonensis LOOC260T 3			
single	30837	Lactobacillus hokkaidonensis LOOC260T 2			
single	30836	Lactobacillus hokkaidonensis LOOC260T 1			

Assembly commands

hgap

Set optional parameters for HGAP pipeline

Select UGE-node to run :

month\_fat (32 CPUs and 320GB memory)

month\_medium (32 CPUs and 256GB memory)

Same results will be generated with either option.  
You can check the CPU and memory usage at [NIG-SC Website](#).

1 : The approximate genome size, in base pairs.(Must be a value between 1 and 150000000)

GenomeSize =

2 : The minimum length of reads (in base pairs) to use as seeds for pre-assembly.

Minimum Seed Length :

Automatic Estimation

If the coverage exceeds 25X, the Minimum Seed Read Length that results in at least 25X coverage by the longest subreads will be calculated automatically. If the coverage is less than 25X, the user-specified value will be used.

Use Manually Specified Value (regardless of the coverage)

BACK RUN

1.メールアドレスを設定

2.実行内容を確認

3.RUNをクリック

4.OKをクリック

Do you really want to execute pipeline programs?

キャンセル

OK

5.STATUSをクリック

The reservation was completed.

STATUS NEXT JOB

**ACCOUNT**

login ID [koshu40]

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**

Running Status

**Status - de novo Assembly**

Mapping Job

Order

Sort by : ID Descending  Show Only Your Own Job

Delete

ID	UserID	Submission accession	P/S	Status	Tool	Read #	Read length	Assembly detail	Mapping detail	Start time	End time
30562	---	---	S	complete	Canu	1,933	---			2018-03-20 09:57:53	00:09:49
30561	---	---								2018-03-20 10:07:42	
30560	---	---									
30557	---	---									
30556	---	---									

2. Show Only Your Own Job をチェック

3. Reload をクリック

1. JOB STATUS の step1. de novo Assembly をクリック

**Status - de novo Assembly**

Mapping Job **de novo Assembly Job** Preprocessing Job

Order

Sort by : ID Descending  Show Only Your Own Job

Delete

ID	UserID	Submission accession	P/S	Status	Tool	Read #	Read length	Assembly detail	Mapping detail	Start time	End time	Elapsed time
30527	koshu40	---	S	complete	HGAP		---			2018-03-19 19:32:28	2018-03-19 22:59:11	03:26:43
30508	koshu40	---	S	complete	HGAP		---			2018-03-19 18:02:33		15:24:14
30430	koshu40	---	S	complete	HGAP		---					175:34:18

3. View をクリック

Detail view

BACK

Job info

<b>ID</b>	30527
<b>Tool (Version)</b>	HGAP (Protocol3(v 2.2.0))

解析ツールとバージョン

クエリファイル

RunAccession or Filename	Download	Read length	Alias
m130821_065825_42195_c100539522550000001823089611241356_s1_p0.3.bax.h5	<a href="#">m130821_065825_42195_c100539522550000001823089611241356_s1_p0.3.bax.h5</a>	N.A. bp	Lactobacillus hokkaidonensis LOOC260T 3
m130821_065825_42195_c100539522550000001823089611241356_s1_p0.2.bax.h5	<a href="#">m130821_065825_42195_c100539522550000001823089611241356_s1_p0.2.bax.h5</a>	N.A. bp	Lactobacillus hokkaidonensis LOOC260T 2
m130821_065825_42195_c100539522550000001823089611241356_s1_p0.1.bax.h5	<a href="#">m130821_065825_42195_c100539522550000001823089611241356_s1_p0.1.bax.h5</a>	N.A. bp	Lactobacillus hokkaidonensis LOOC260T 1

Download modified queries

The modified query file does not exist, because of the following reasons.

- The file is expired. (about 1 months)
- Job is waiting for execution queue.
- Error in query file.

DDBJ 登録用  
ファイル

Download wgs file

- [out\\_WGS.fasta.gz](#) (Original size 2.4 MB)

アセンブリ統計値

Assembly statistics

Contig # : 4  
 Total contig size : 2,433,614  
 Maximum contig size : 2,289,497  
 Minimum contig size : 11,372  
 N50 contig size : 2,289,497

実行時間

Time

Wait time	Start time	End time
3days 1: 38:48	2018-03-19 19:32:28	2018-03-19 22:59:11

実行結果

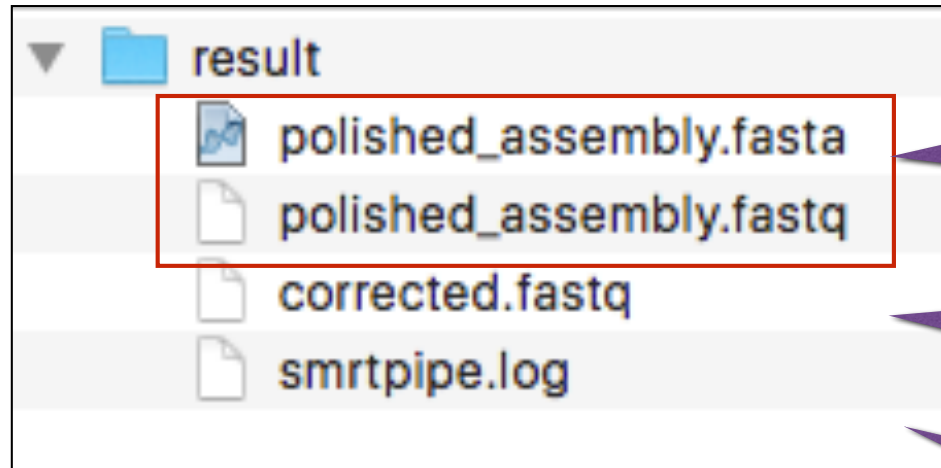
Command	Start time	End time	Log1	Log2	Result	MD5
run HGAP through smrtpipe.py : GenomeSize=2500000,minSeedLength=6000	2018-03-19 19:32:28	2018-03-19 22:58:07	<a href="#">View</a>		<a href="#">Download(13.1 MB)</a>	<a href="#">MD5</a>

1.Download を  
クリック

BACK

# 結果ファイル

---



アセンブル結果ファイル

補正された ロングリード

ログファイル

**ご清聴ありがとうございました**