パイプライン

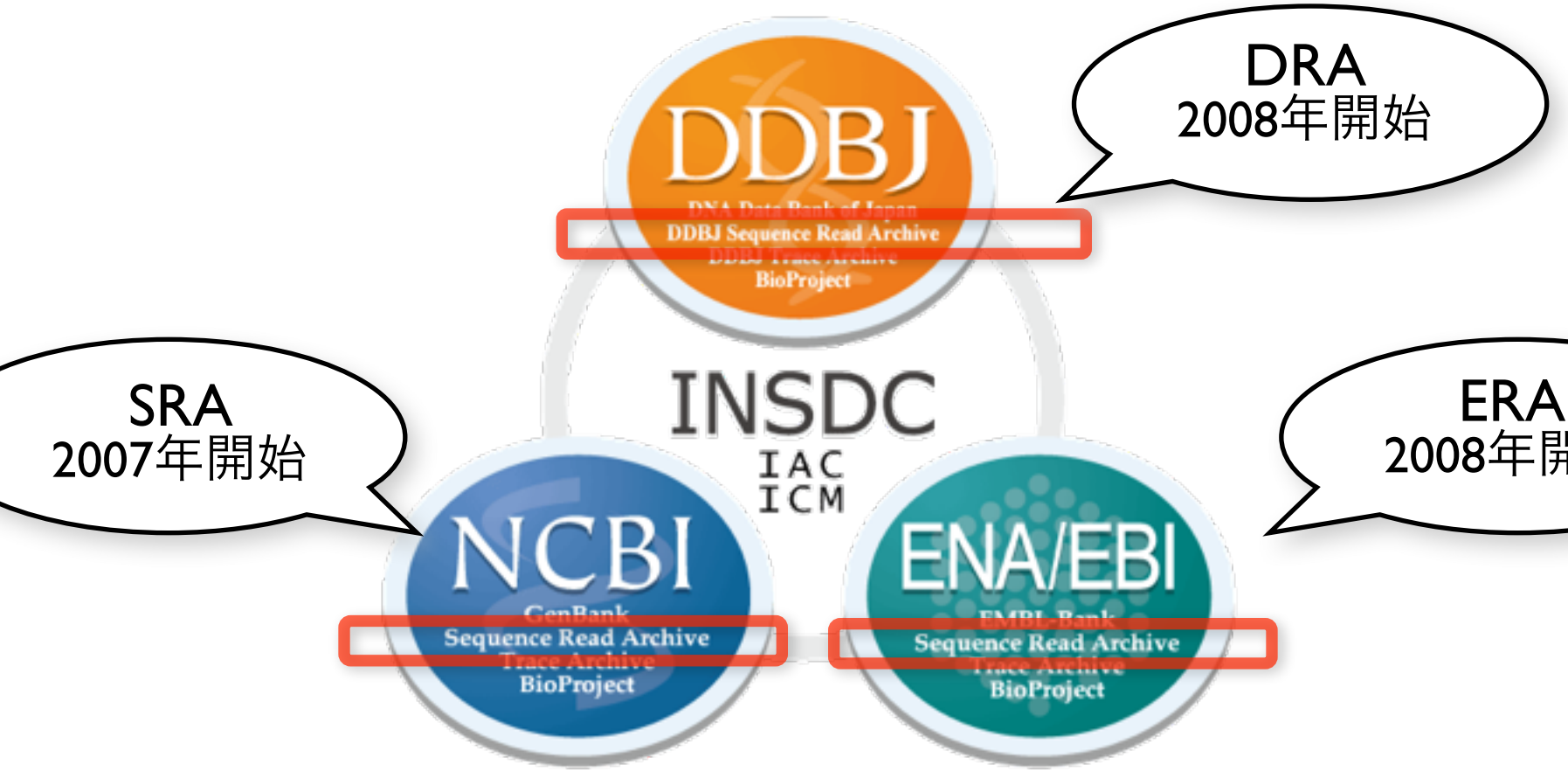
遺伝研大量研・DDBJセンター
中村保一

DRA

DDBJ Sequence Read Archive

DDBJ Sequence Read Archive (DRA)

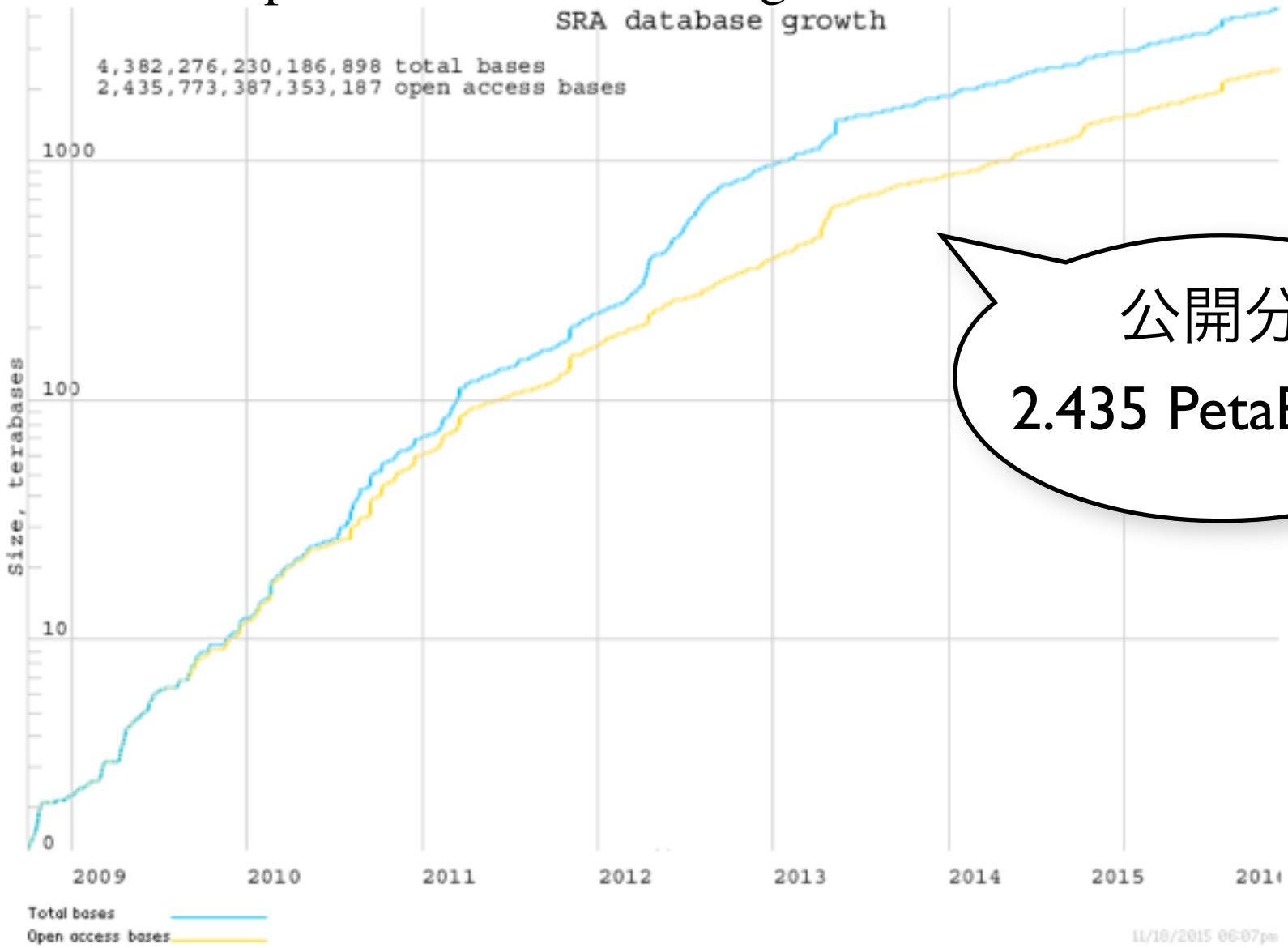
新世代シークエンサから出力される配列や
アライメントデータを登録・公開



International Nucleotide Sequence Databank Collaboration

SRA growth (NCBI)

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



公開分

2.435 PetaBases

DRAウェブサイト ⇒ [DRA] で検索

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

公開データの DRA Search での検索

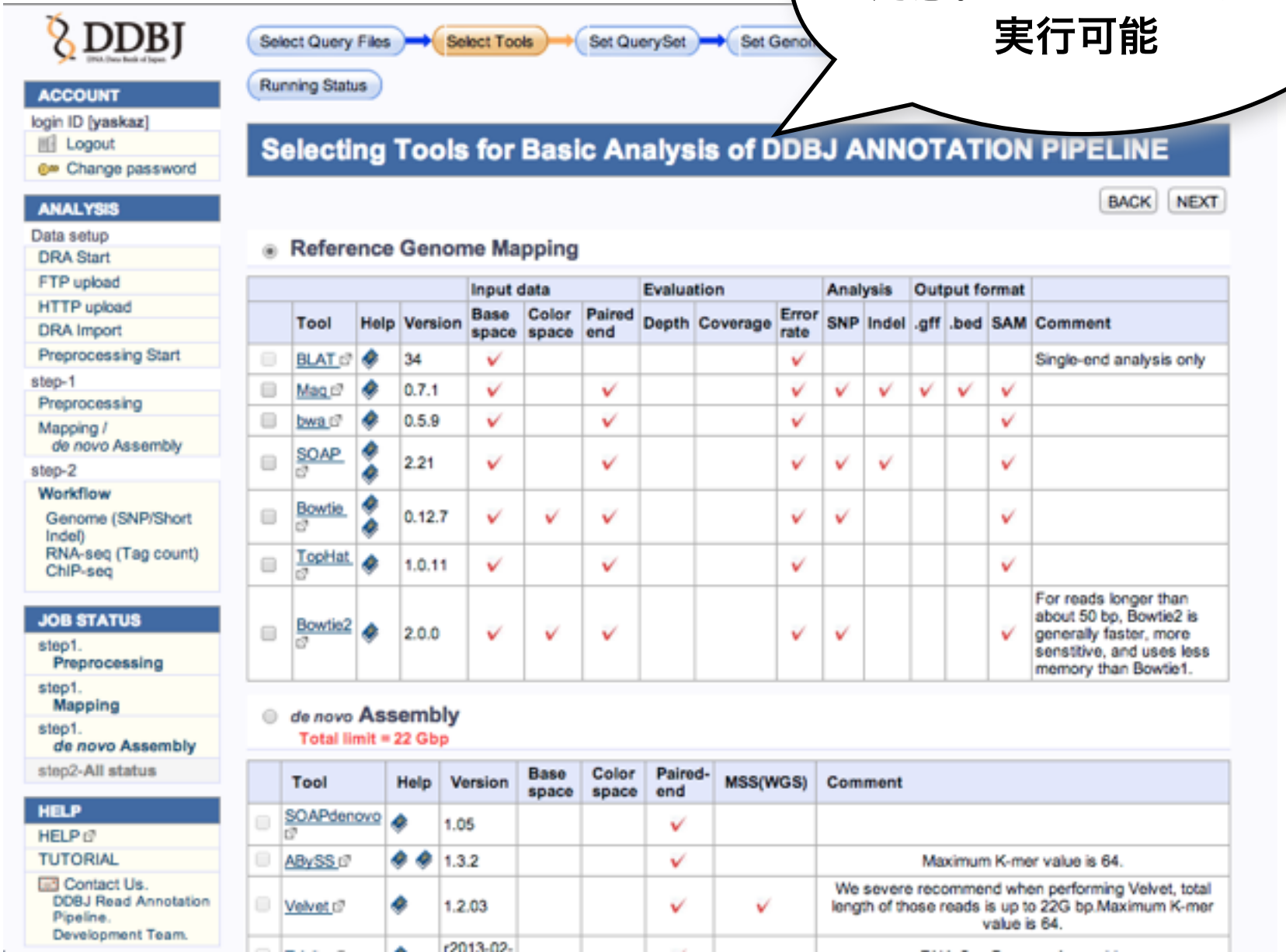
公開データは EBI SRA / NCBI SRA と共有されています

解析パイプラインも提供してます

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

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

Select Query Files → Select Tools → Set QuerySet → Set Genor

Running Status

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

				Input data			Evaluation			Analysis		Output format			
	Tool	Help	Version	Base space	Color space	Paired end	Depth	Coverage	Error rate	SNP	Indel	.gff	.bed	SAM	Comment
	BLAT		34	✓					✓						Single-end analysis only
	Maq		0.7.1	✓		✓			✓	✓	✓	✓	✓	✓	
	bwa		0.5.9	✓		✓			✓					✓	
	SOAP		2.21	✓		✓			✓	✓	✓			✓	
	Bowtie		0.12.7	✓	✓	✓			✓	✓				✓	
	TopHat		1.0.11	✓		✓			✓					✓	
	Bowtie2		2.0.0	✓	✓	✓			✓	✓				✓	For reads longer than about 50 bp, Bowtie2 is generally faster, more senstitive, and uses less memory than Bowtie1.

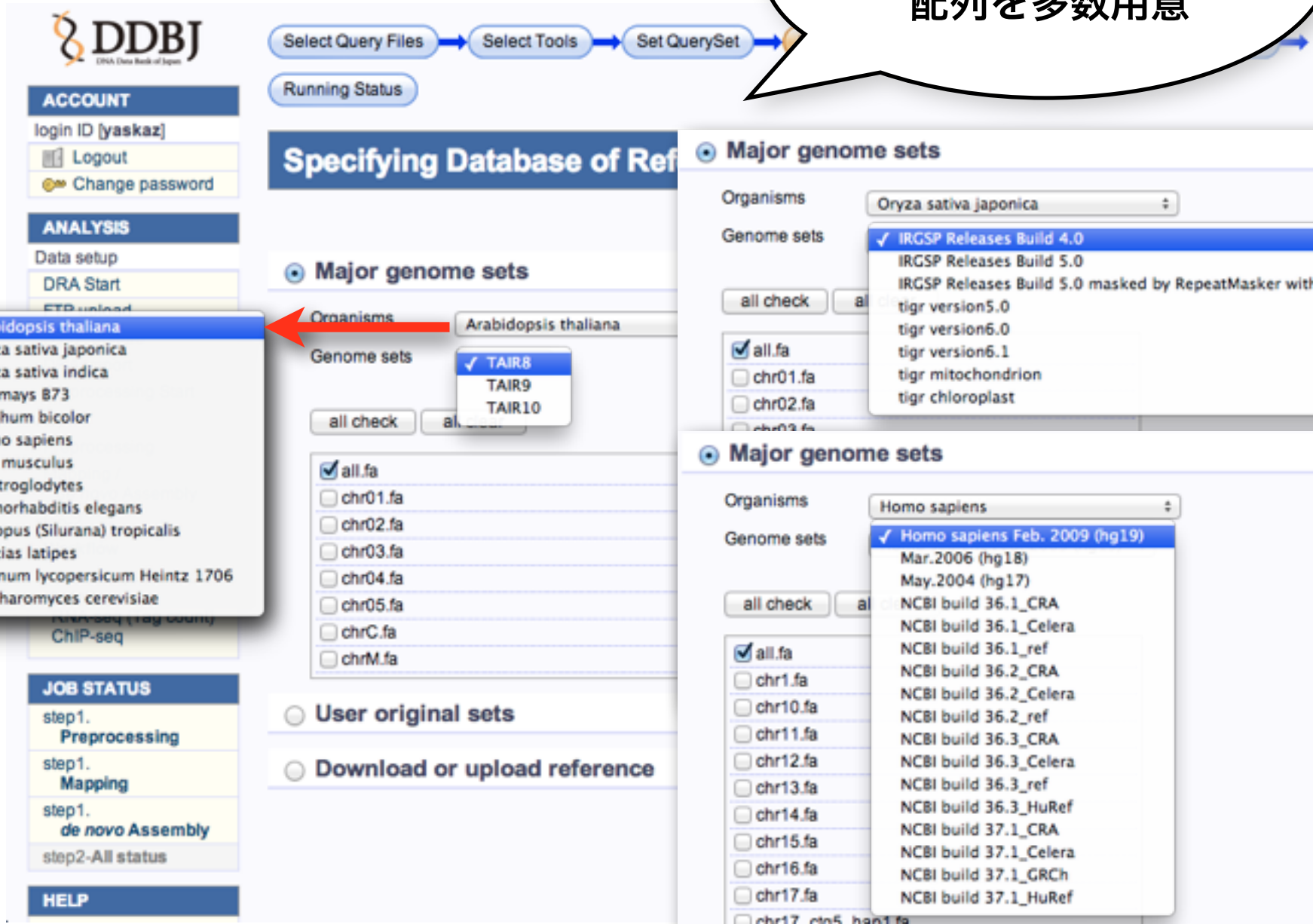
de novo Assembly

Total limit = 22 Gbp

	Tool	Help	Version	Base space	Color space	Paired-end	MSS(WGS)	Comment
	SOAPdenovo		1.05			✓		
	ABySS		1.3.2			✓		Maximum K-mer value is 64.
	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.
			r2013-02-					

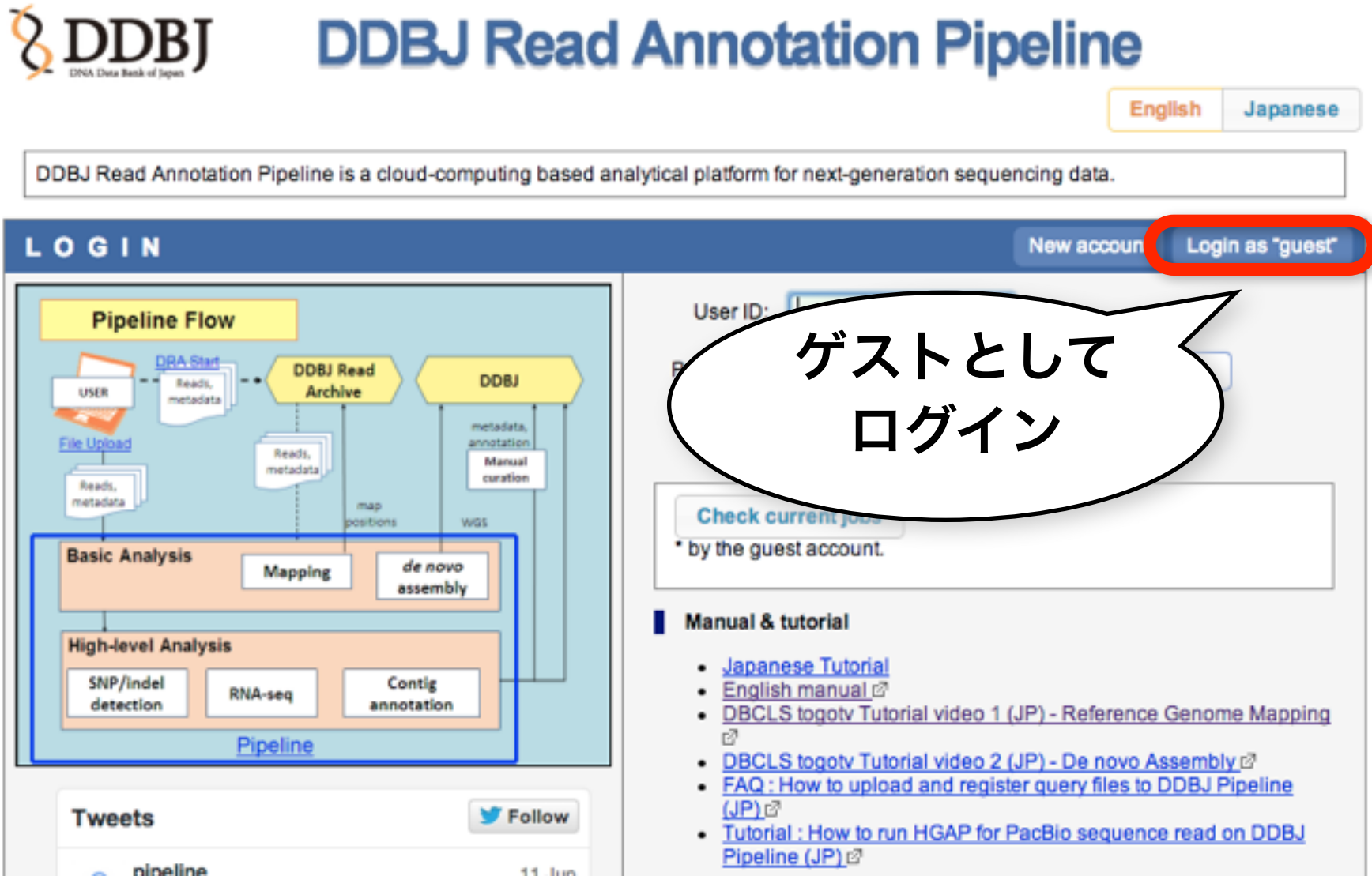
DDBJ pipeline: 比較対象

イネ、マウスなど
解析比較対象となる
配列を多数用意

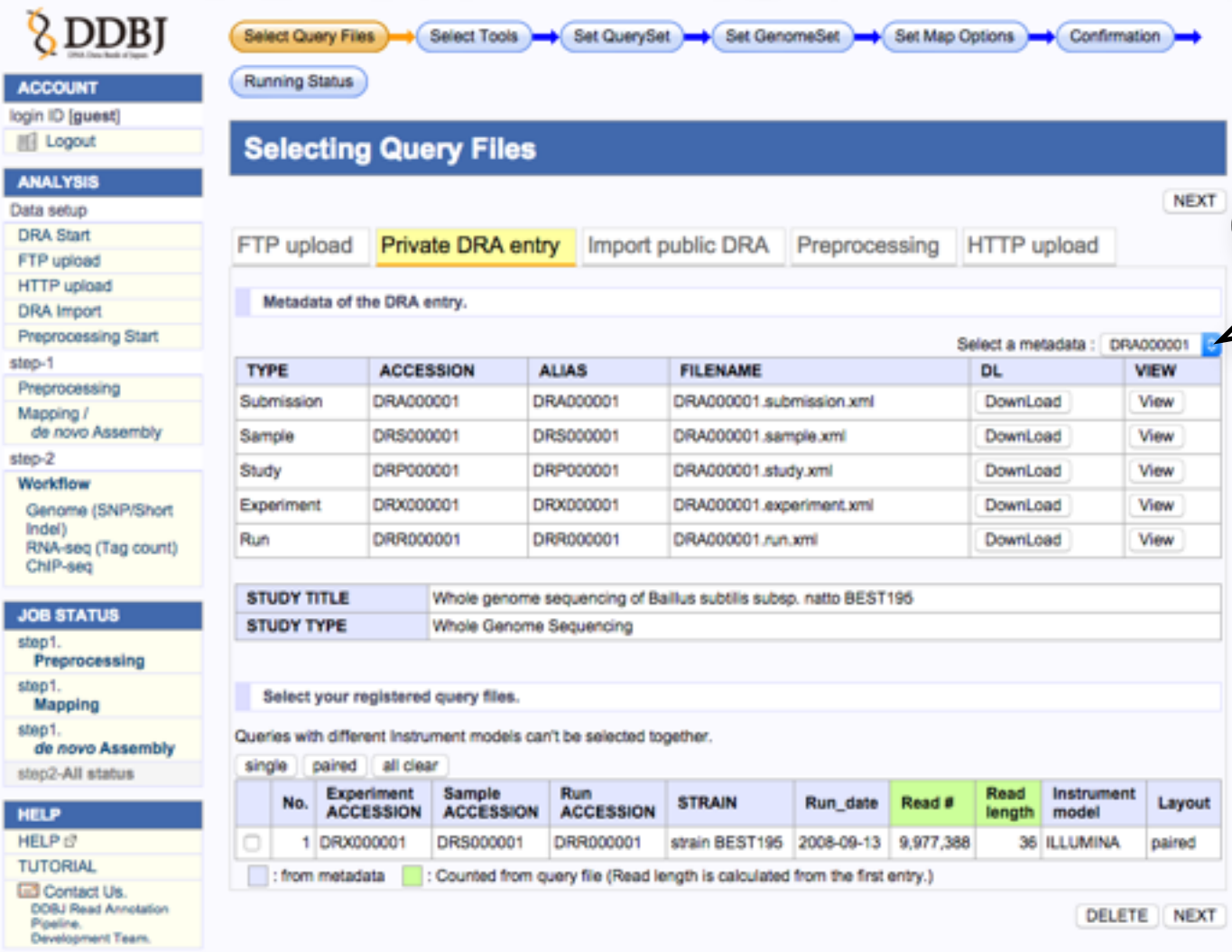


DDBJ パイプライン、体験してみましょう

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



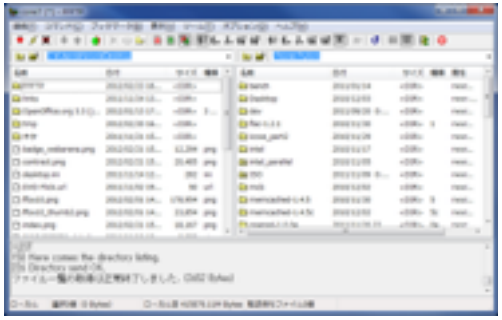
処理に使うNGSの配列ファイルの用意



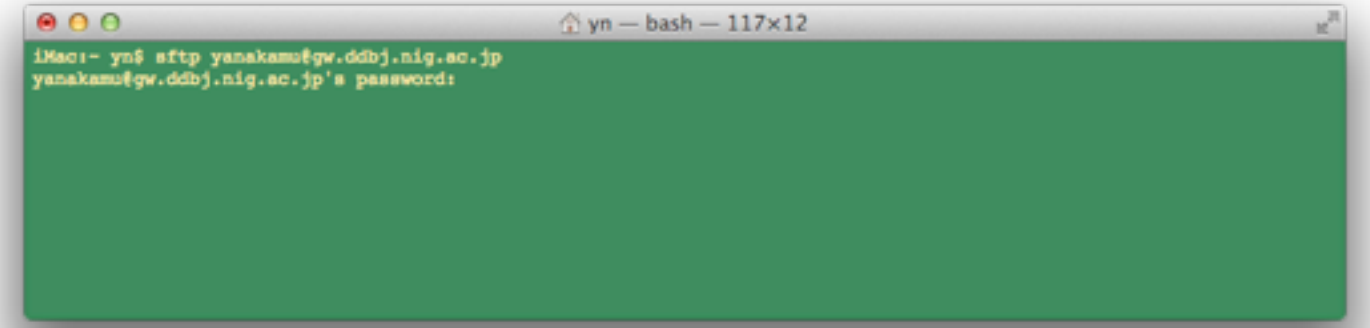
処理に使うNGSの配列ファイルの用意

FTP (file transfer protocol) クライアント

Cyberduck

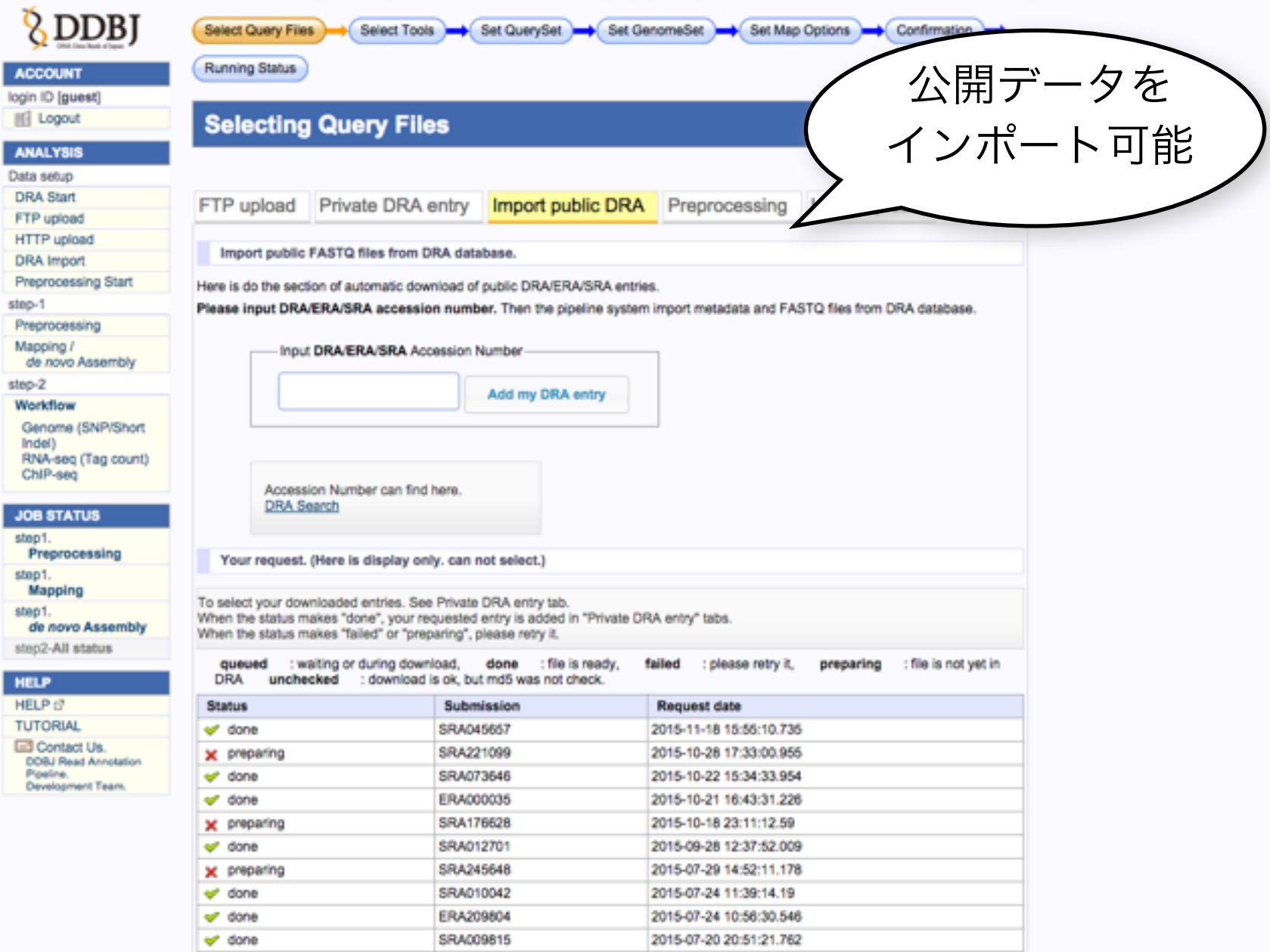


FFFTP

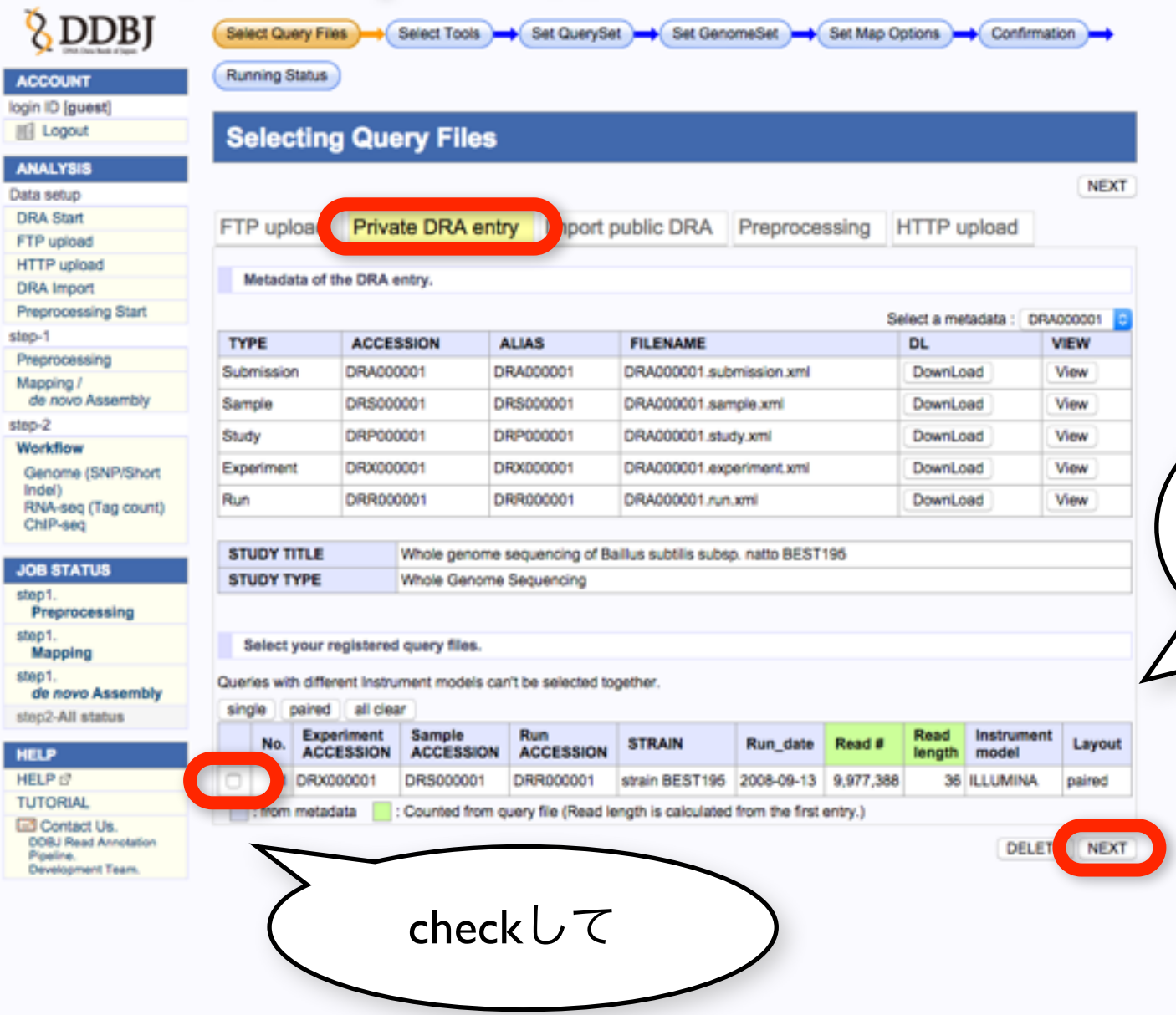


Cyberduck のような専用クライアントを使えばドラッグ＆ドロップでファイル転送可能。コマンドでの送受信もたいして難しくありませんが。

処理に使うNGSの配列ファイルの用意



今回はupload済のエントリから



納豆菌の
公開データが
インポート済

checkして

DDBJ

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

ACCOUNT
login ID [guest]
Logout

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

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

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

Selecting Tools for Basic Analysis of DDBJ ANNOTATION PIPELINE

BACK NEXT

Reference Genome Mapping

	Tool	Help	Version	Input data: Base space	Input data: Color space	Input data: Paired end	Evaluation: Depth	Evaluation: Coverage	Evaluation: Error rate	Analysis: SNP	Analysis: Indel	Output format: .gff	Output format: .bed	Output format: SAM	Comment
☐	BLAT		34	✓					✓						Single-end analysis only
☐	bwa		0.6.1	✓		✓	✓	✓	✓					✓	
☐	Bowtie		0.12.7	✓	✓	✓	✓	✓	✓	✓				✓	
☐	TopHat		1.0.11	✓		✓	✓	✓	✓					✓	
☐	Bowtie2		2.0.0	✓	✓	✓	✓	✓	✓	✓				✓	For reads longer than about 50 bp, Bowtie2 is generally faster, more senstitive, and uses less memory than Bowtie1.
☐	TopHat2		2.0.9	✓		✓	✓	✓	✓					✓	

de novo Assembly

Total limit = 22 Gbp

	Tool	Help	Version	Base space	Color space	Paired-end	MSS(WGS)	Comment
☐	SOAPdenovo		1.05	✓		✓		
☐	ABySS		1.3.2	✓		✓		Maximum K-mer value is 64.
☑	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.
☐	Trinity		r2013-02-25	✓		✓		RNA-Seq De novo Assembly
☐	Platanus		1.2.2	✓		✓		
☐	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.

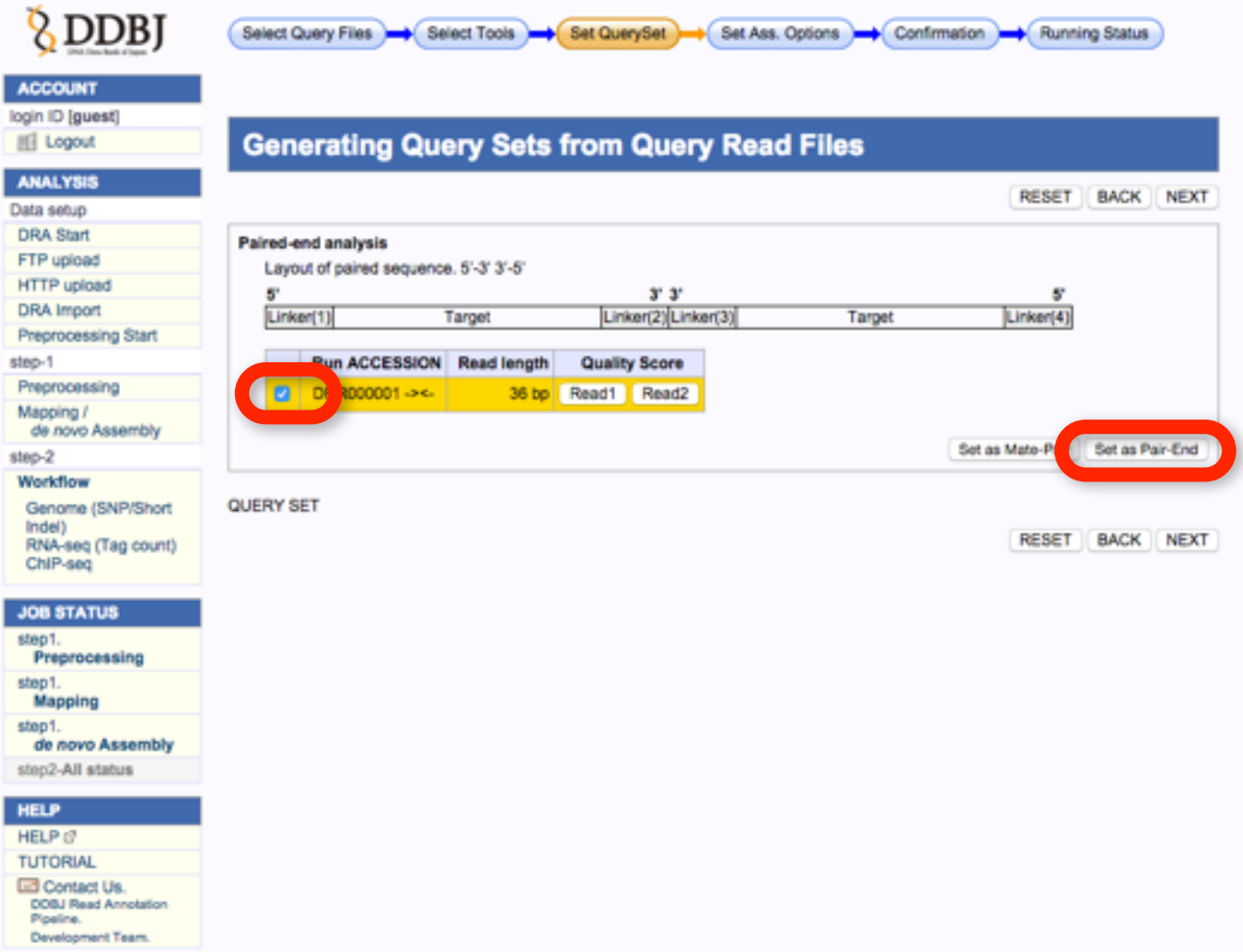
The contigs will be aligned to reference genome.

	Tool	Comment
○	BLAT	Single-end analysis only

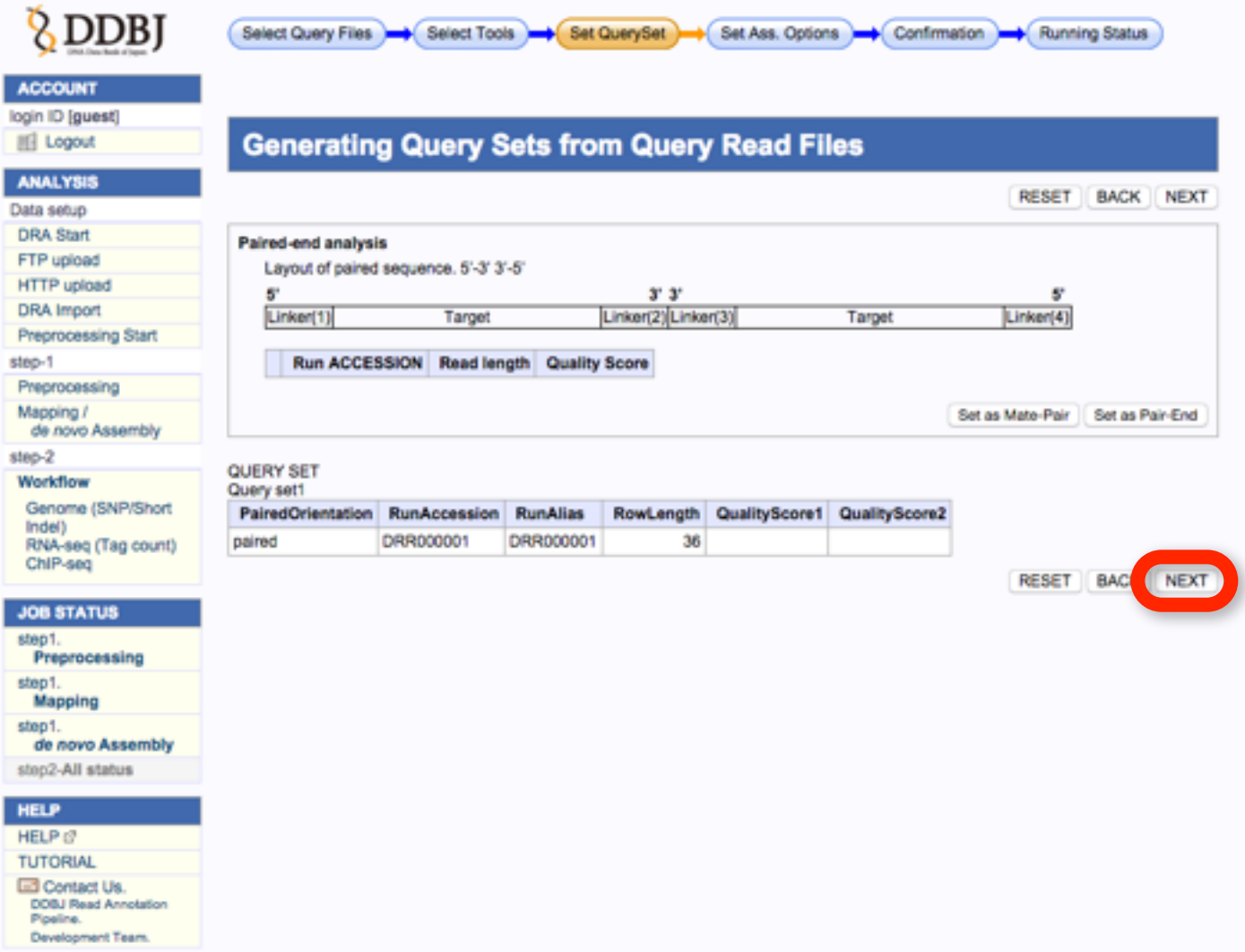
BAC NEXT

velvet で
アセンブル
しましょう

配列とペア形式のセットを選択

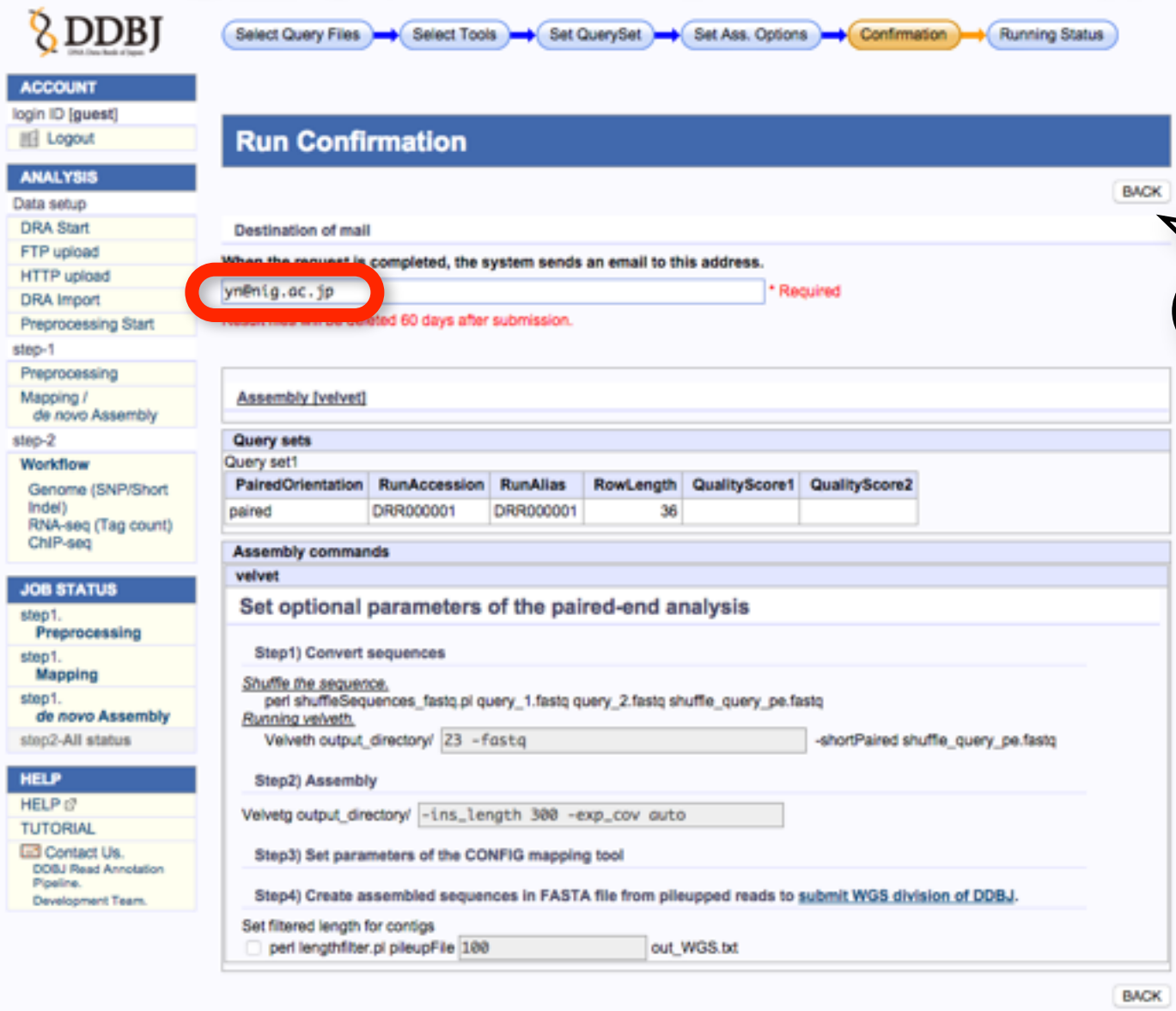


配列のセットの形式を選んで次へ



オプションのパラメータを選びます

連絡先をいれたら、RUN ボタン押すだけ！



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

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

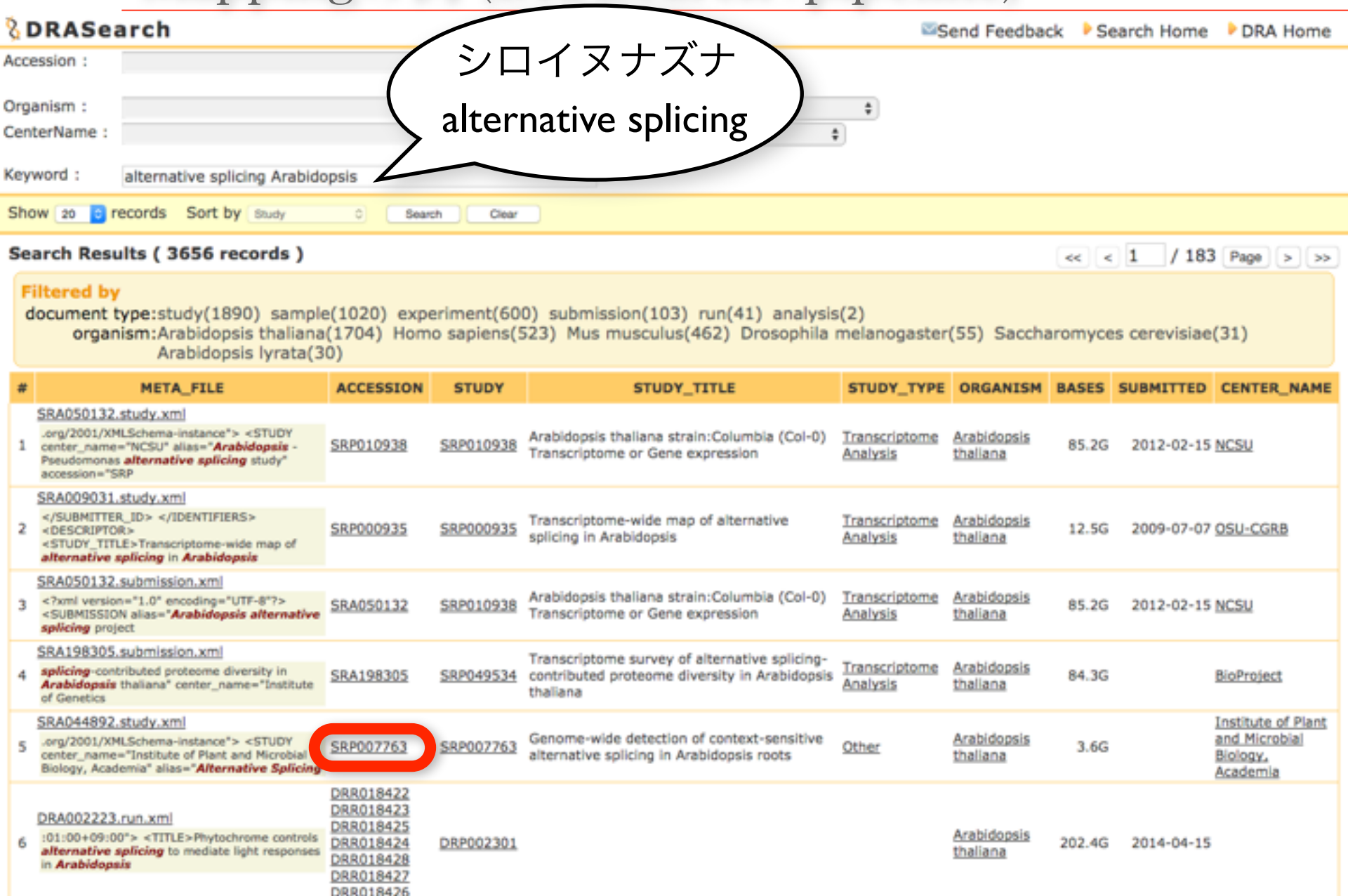
Wait time	Start time	End time
0: 1:22	2013-01-11 22:06:40	2013-01-11 22:13:40

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

BACK

結果ファイル

Mappingの例 (DRAsearch+pipeline)



#	META_FILE	ACCESSION	STUDY	STUDY_TITLE	STUDY_TYPE	ORGANISM	BASES	SUBMITTED	CENTER_NAME
1	SRA050132.study.xml .org/2001/XMLSchema-instance"> <STUDY center_name="NCSU" alias="*Arabidopsis* - Pseudomonas *alternative splicing* study" accession="SRP	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 </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	SRA050132.submission.xml <?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	SRA198305.submission.xml *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	SRA044892.study.xml .org/2001/XMLSchema-instance"> <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	DRA002223.run.xml :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

DRASearch

Send Feedback | Search Home | DRA Home

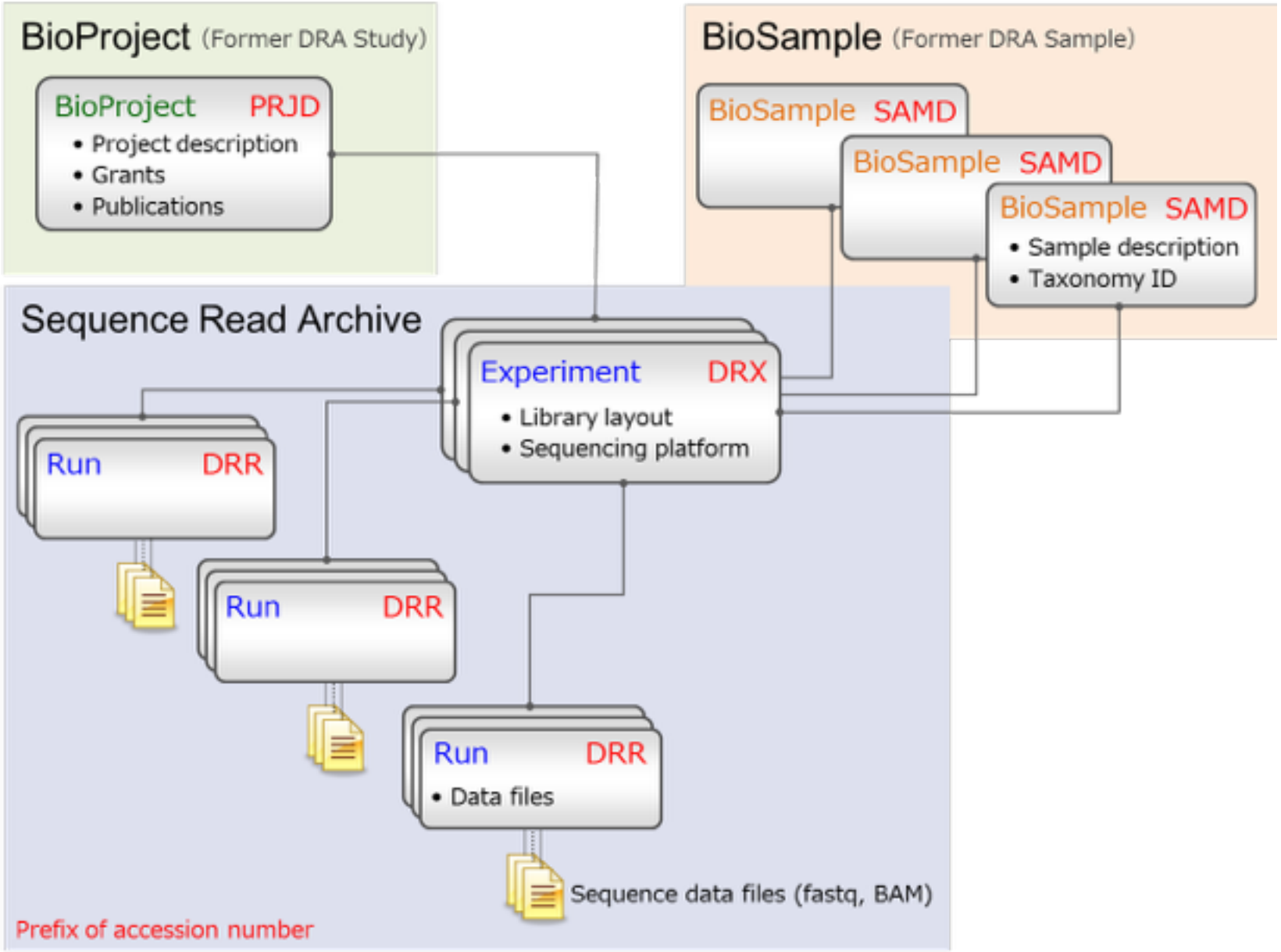
SRP007763

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	
	SRA045009	FTP	
Experiment	[illegible]	FASTQ	SRA
	SRX092048	FASTQ	SRA
	SRX092049	FASTQ	SRA
	SRX092050	FASTQ	SRA
Sample	SRS256250		
	SRS257585		
	SRS257586		
	SRS257587		

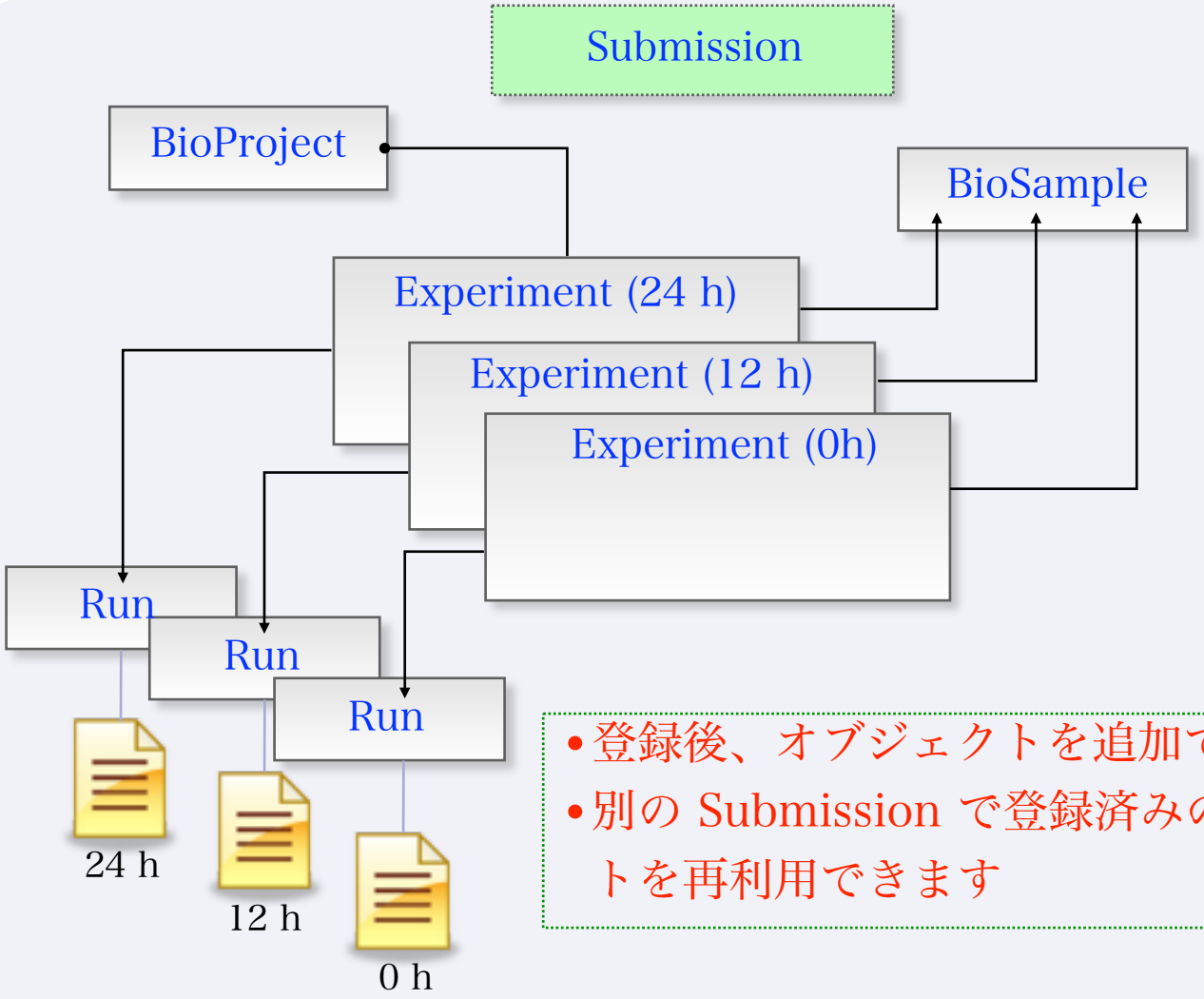
SRA の Metadata

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



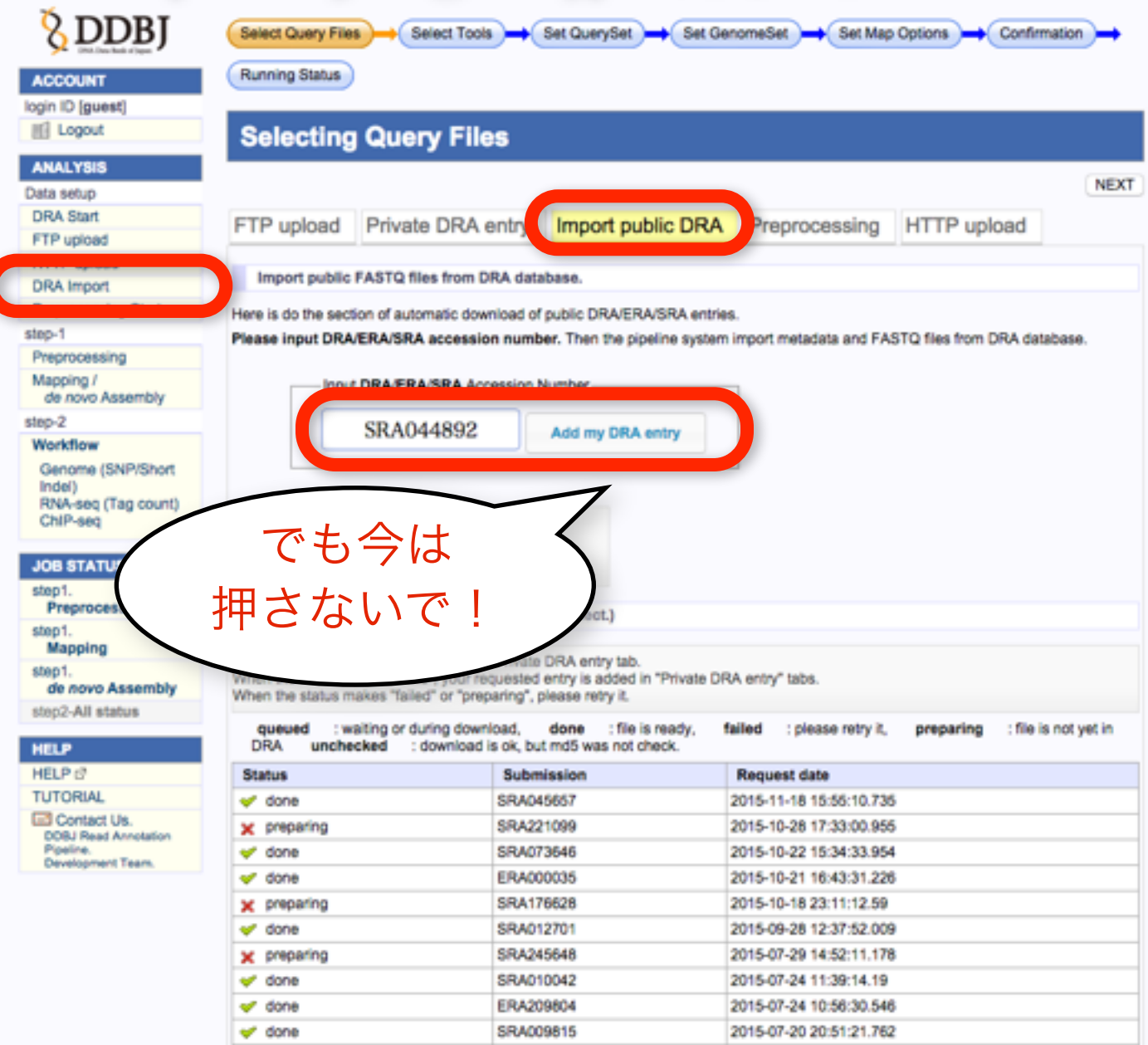
メタデータ構成の例

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

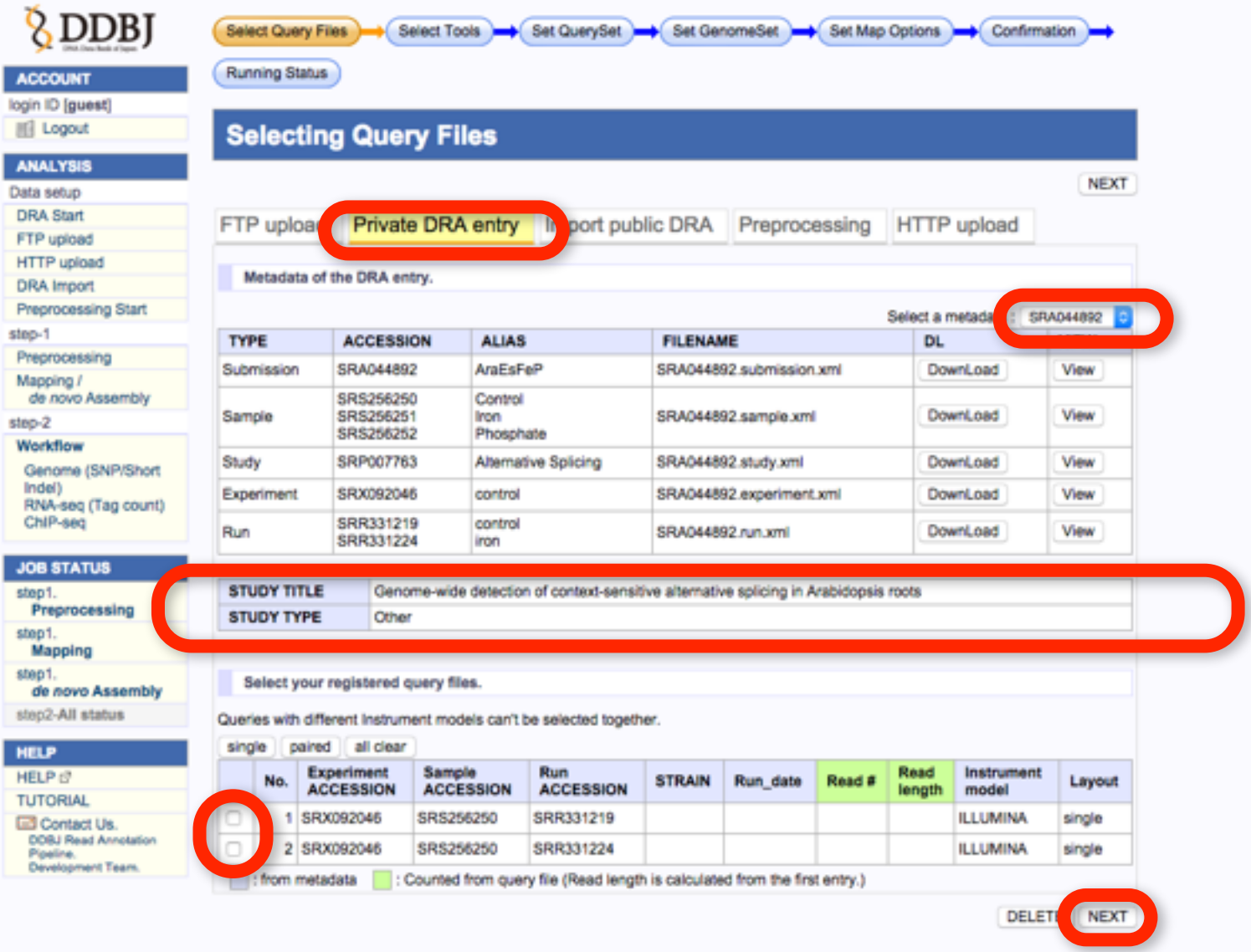


- 登録後、オブジェクトを追加できます
- 別の Submission で登録済みのオブジェクトを再利用できます

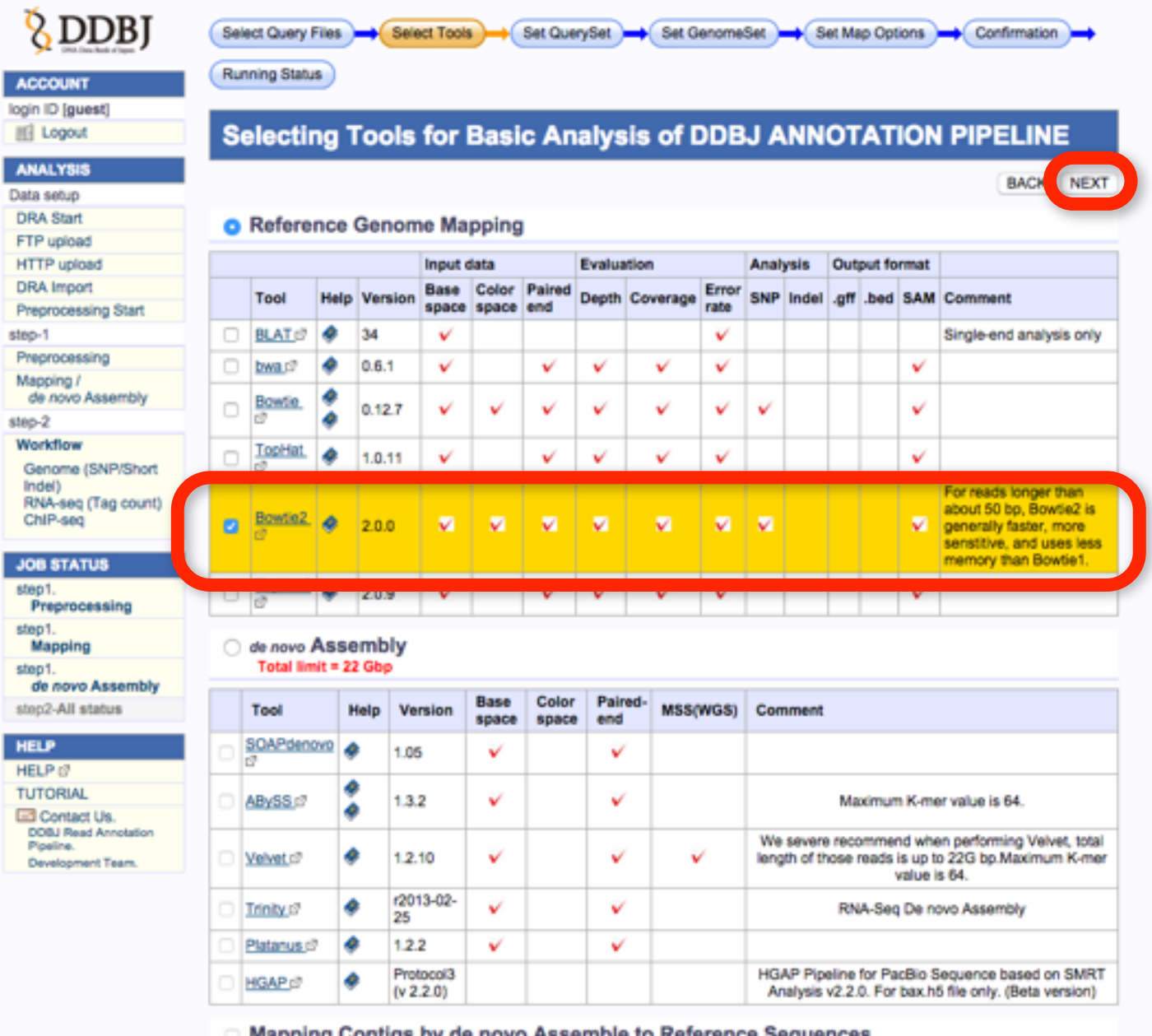
p.ddbj.nig.ac.jp を開き、さっきのIDを入力



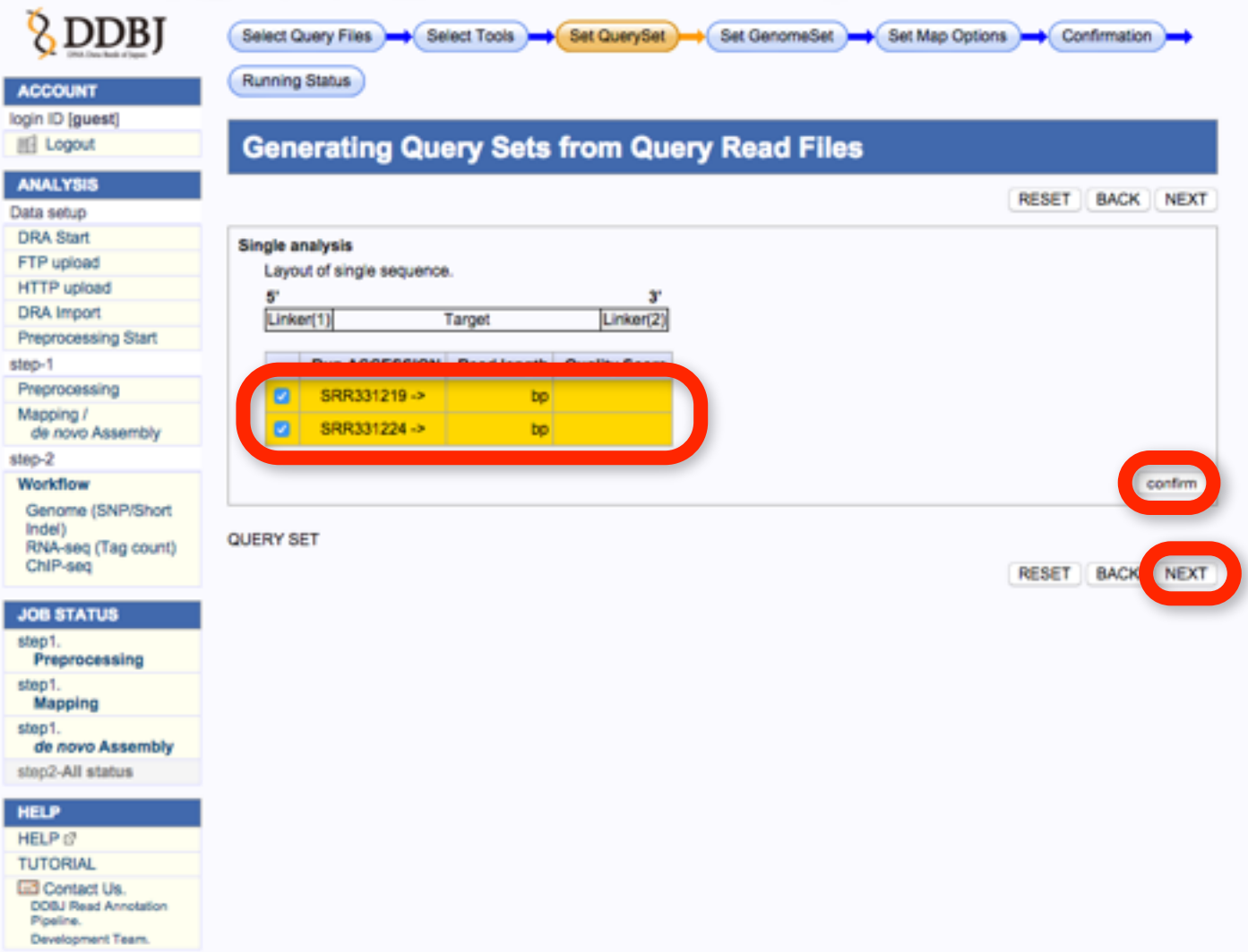
SRA044892 をロードしておきました



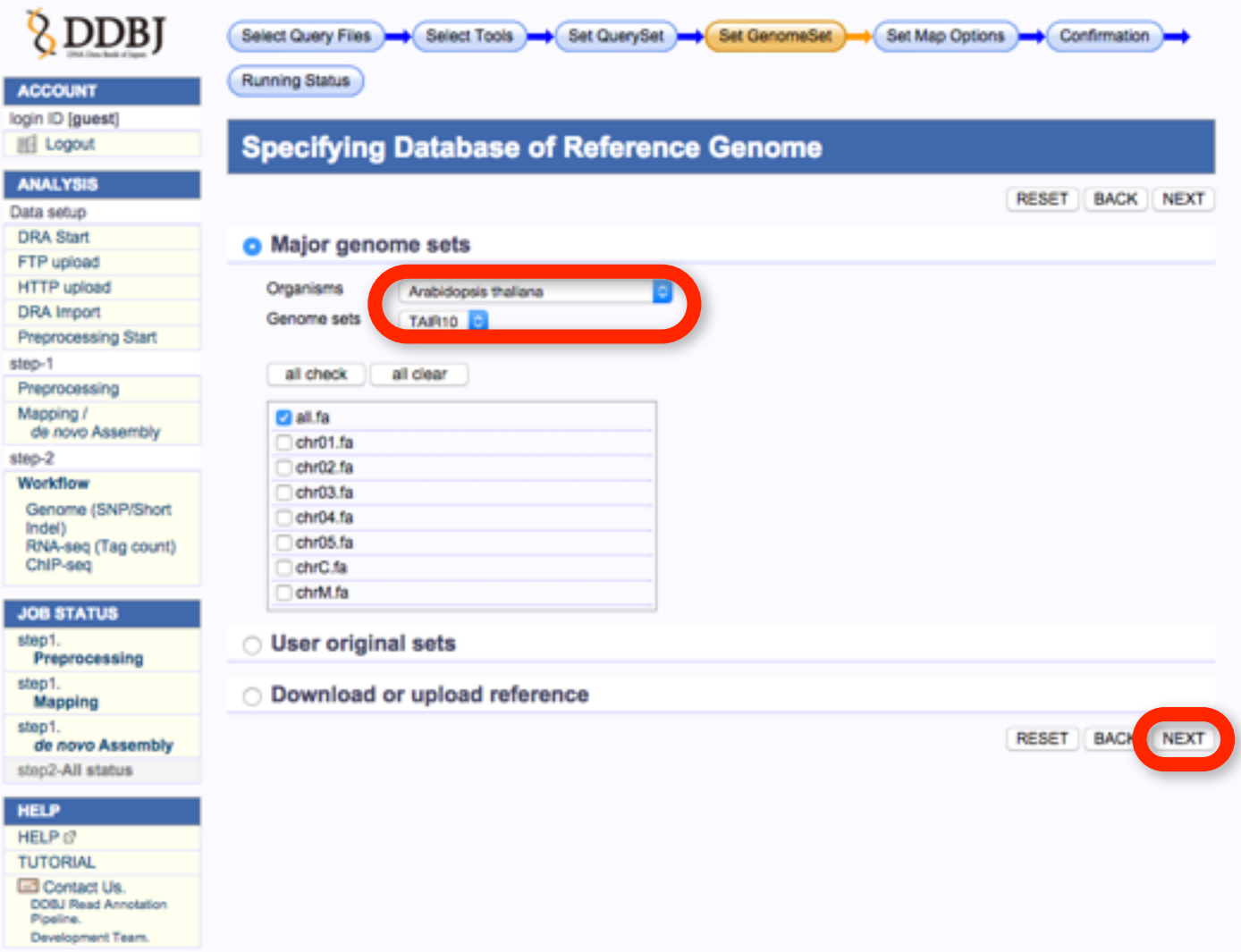
Bowtie2 を選んで NEXT



配列を選んで confirm, NEXT

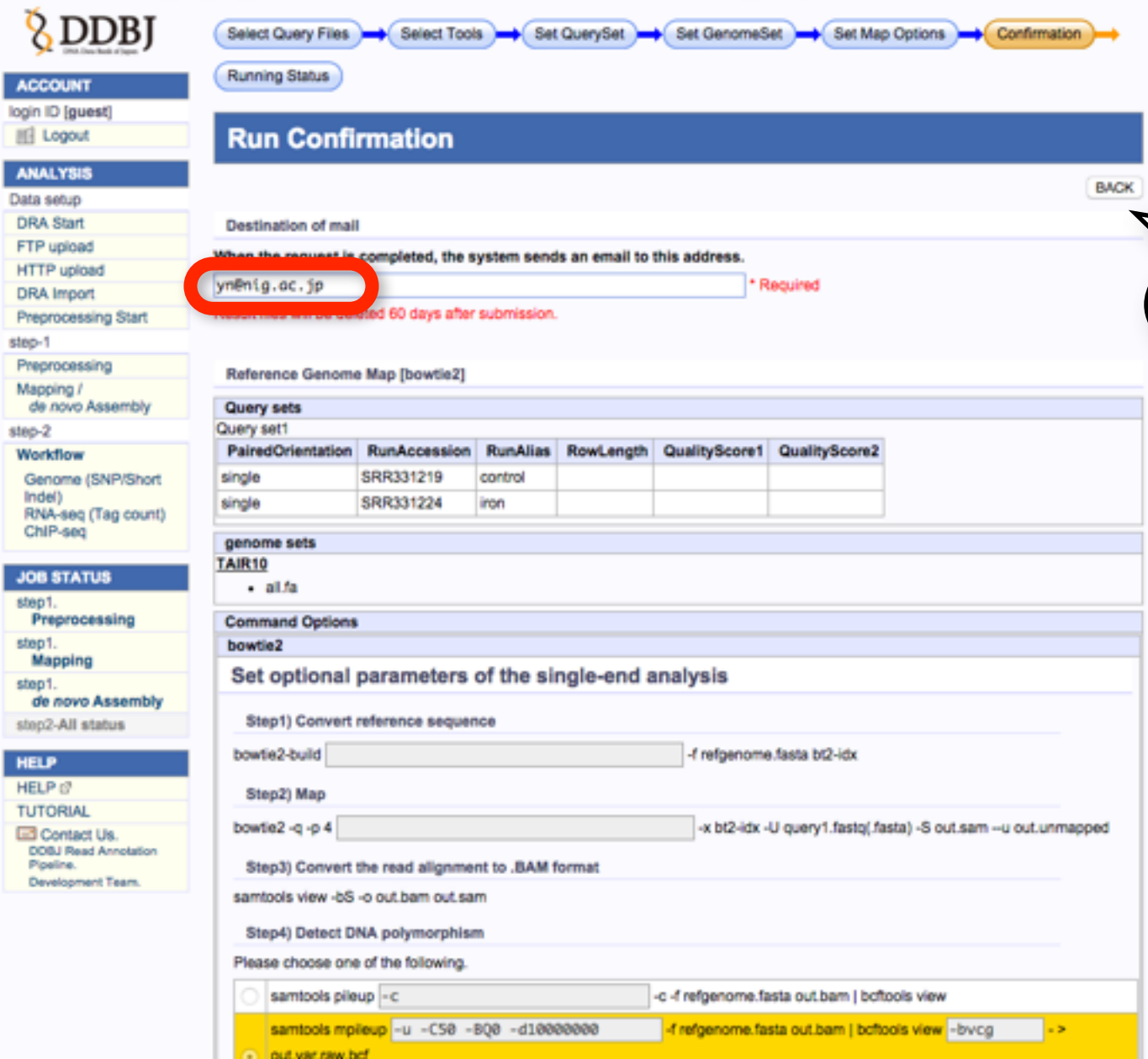


リファレンスに TAIR10 を選んでNEXT



option 変更なければそのままNEXT

終了したらメイルが来ます



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

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

実行結果

Development Team.

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

大学共同利用法人　情報・システム研究機構　国立遺伝学研究所

スーパーコンピュータシステム

SuperComputer Facilities of National Institute of Genetics

サイトポリシー　サイトマップ

検索...　検索

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

システム使用方法

- 基本的利用方法
- その他UGE利用方法
- ファイル転送方法
- システム利用TIPS
- 稼働スケジュール

各種申請

- システムの利用条件
- ログインユーザアカウント発行基準
- 各種申請窓口について
- 新規ユーザ登録申請
- SSH公開鍵登録
- リソース拡張申請
- アカウント継続・停止申請
- 登録内容変更申請
- OSSインストール申請

Webサービス

- MiGAP利用申請
- MiGAPパスワード変更申請
- DDBJ Pipeline利用申請

重要なお知らせ

一覧へ

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	【UGE障害】スーパーコンピュータシステム UGE動作不良のお知らせ
2014年9月10日	【スパコンユーザ会】会議報告
2014年3月6日	2014年3月6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用拠点として 最新鋭の大規模クラスタ型計算機、大規模メモリ共有型計算機、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ティング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。DDBJセンターとして提供する各種サービスについてはDDBJ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

lustreのsize,file項目は、現在のディスク使用量（全ユーザ合計）／ディスク容量で表現しています。
quotaのsize項目は、申請ディスク使用量（全ユーザ合計）／ディスク容量で表現しています。
sizeの単位は、Tbyteです。

lustre1 size :	963.76 /	1,024.90	94%	lustre1 file :	197,742,325 /	731,840,512	27%
quota size :	2,033.08 /	1,024.90	198%				
lustre2 size :	921.38 /	1,024.90	89%	lustre2 file :	685,686,521 /	731,840,512	93%
quota size :	1,143.66 /	1,024.90	111%				
lustre3 size :	[illegible] /					731,840,512	23%
quota size :							
lust[illegible]							4%
quo[illegible]							
lust[illegible]							57%
quota size :							

こちらからどうぞ！

[illegible] 利用状況概要

現在のUGE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。