DRA と DDBJパイプライン入門

中村 保一

Aug 24, 2017

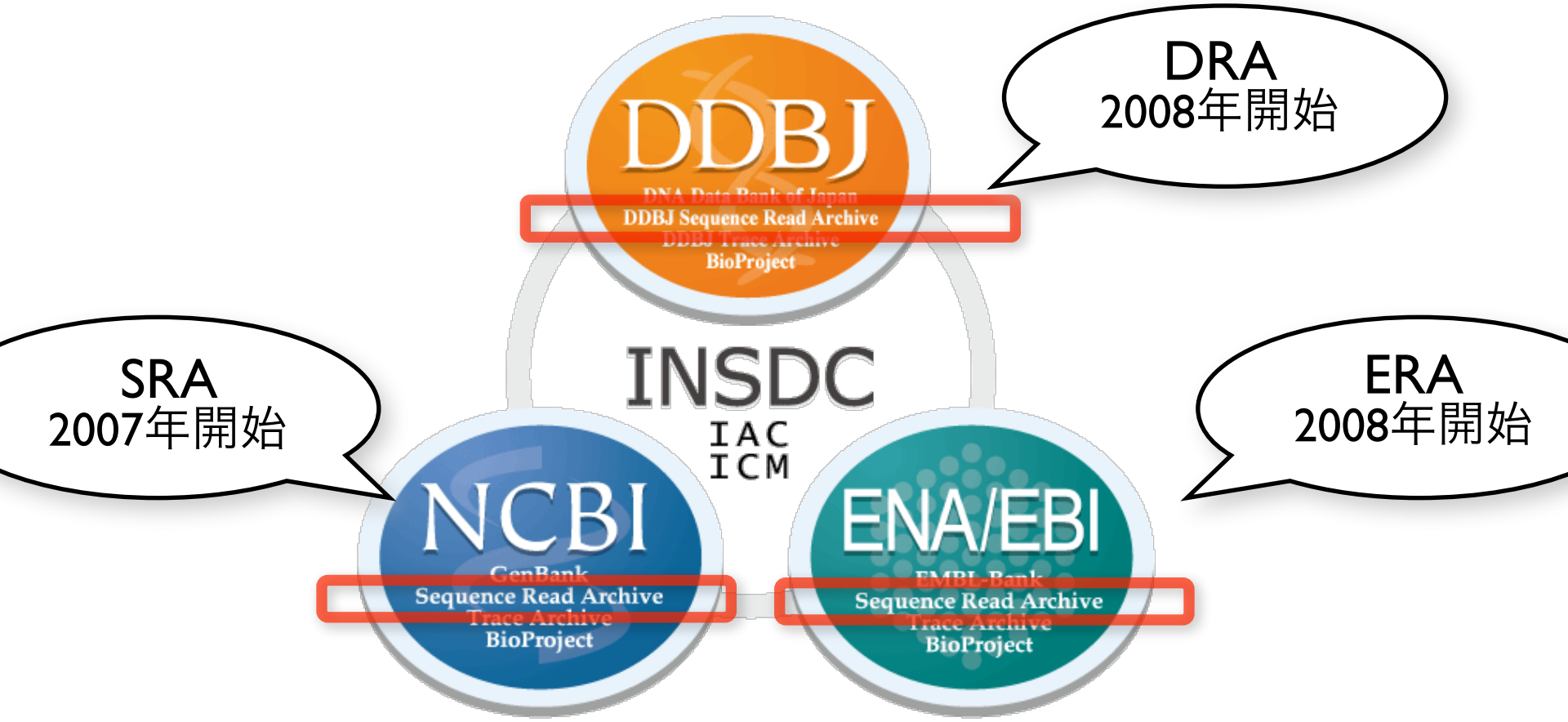
石川県立大学

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

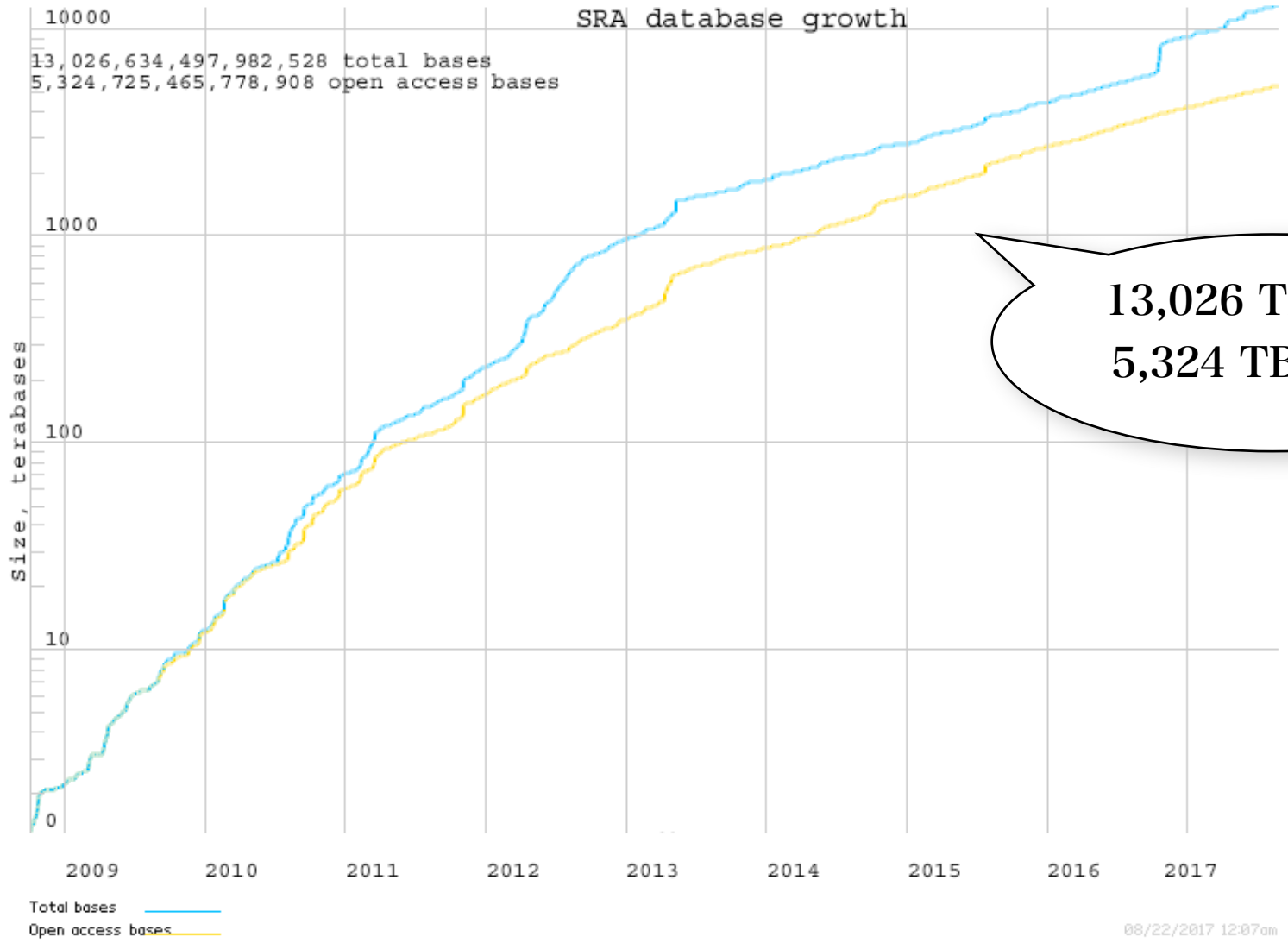


International Nucleotide Sequence Databank Collaboration

SRA growth (NGS row/bam data amount)

DDBJ DNA Data Bank of Japan

NCBI Sequence Read Archive



13,026 TB total
5,324 TB open

DRAウェブサイト ⇒ [DRA] で検索

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

DRA: 新型シーケンサデータを保存・共有

DDBJ
DNA Data Bank of Japan

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 DRA

DDBJ Sequence Read Archive (DRA) は Roche 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD® S
などの次世代シークエンサからの出力データのためのデータベースです。 DRA は International Nucleotide Sequence Database
Collaboration (INSDC) のメンバーであり， NCBI Sequence Read Archive (SRA) と EBI Sequence Read Archive (ERA) との国際
もと，運営されています。 従来のキャピラリ式シークエンサからの出力データは DDBJ Trace Archive にご登録ください。

検索
データをキーワード，生物名，シークエンサなどで検索する

登録
新型シークエ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みる

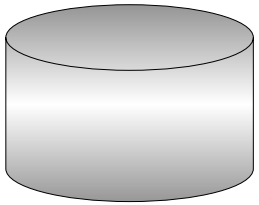
Released Data

search result : 30

Accession	Study Title	Organism(s)	Center Name	Release Date
Reset	Reset	Reset	Reset	Reset
DRA000001	Whole genome sequencing of Baillus subtilis subsp. natto BEST195	Bacillus subtilis subsp. natto	KEIO	2010-03-26
DRA000002	Whole genome resequencing of Bacillus subtilis subsp. subtilis str. 168	Bacillus subtilis subsp. subtilis str. 168	KEIO	2010-03-26
DRA000010	Whole genome shotgun sequences of Oryza sativa japonica variety, Koshihikari	Oryza sativa Japonica Group	NIAS	2010-03-31
DRA000030	Whole-genome DNA methylation analysis in human breast cancer cell lines using MeDIP-seq	Homo sapiens	KUGSPS	2010-03-01
DRA000039	genetic variation detected in 206 klebsiella pneumoniae plasmids	Klebsiella pneumoniae	WMC	2009-12-14
DRA000067	B. anthracis BA103 genome analysis	Bacillus anthracis	NIID	2010-04-22
DRA000068	B. anthracis BA104 genome analysis	Bacillus anthracis	NIID	2010-04-22
DRA000069	Whole SNPs analysis of ciprofloxacin resistance among B. anthracis strains	Bacillus anthracis	NIID	2010-04-22
DRA000070	Whole SNPs analysis of ciprofloxacin resistance among B. anthracis strains	Bacillus anthracis	NIID	2010-04-22
DRA000155	CAGE analysis of whole adult brain and whole embryo rat transcriptome	Rattus norvegicus	RIKEN_OSC	2010-03-17
DRA000169	Linking new promoters to functional transcripts in small samples with nanoCAGE and CAGEscan	Homo Sapiens	RIKEN_OSC	2010-06-08
DRA000205	A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness	Homo Sapiens	RIKEN_OSC	2010-07-23
DRA000220	Whole genome sequencing of Oryzias latipes Hd-rR	Oryzias latipes	KEIO-SM	2010-08-16
SRA002052	Toxoplasma gondii transcript sequencing project	Toxoplasma gondii	UT-MGS	2009-07-01
SRA002053	Glossina morsitans transcript sequencing project	Glossina morsitans	UT-MGS	2009-07-01
SRA002054	Glossina morsitans transcript sequencing project	Glossina morsitans	UT-MGS	2009-06-25
SRA002055	Anopheles stephensi transcript sequencing project	Anopheles stephensi	UT-MGS	2009-07-01
SRA002056	Cryptosporidium parvum transcript sequencing project	Cryptosporidium parvum	UT-MGS	2009-07-01
SRA002057	Plasmodium yoelii transcript sequencing project	Plasmodium yoelii	UT-MGS	2009-09-22

ページが表示されました　インターネット

研究者



インターネットを通じて、登録・公開したデータを、別の研究者が再利用できる

FTP ディレクトリ /ddbj_database/dra/ / ftp.ddbj.nig.ac.jp

この FTP サイトはエクスプローラーでは表示するには、ページ をクリックして、エクスプローラーで FTP サイトを開く をクリックしてください。

```
-Welcome to DDBJ FTP Archive, running on ftp.ddbj.nig.ac.jp!
-
-Please contact ddbj@ddbj.nig.ac.jp when you have any problem for getting
-access to this archive, downloading the data, and etc.

-Termination of DDBJ-XML output format.
-Here is the announcement.
-http://www.ddbj.nig.ac.jp/whatsnew/2010/100625-e.html
-
-A new directory, "patent", was made under "ddbj_database".
-Now, all of patent amino acid sequence data for JPO and KIPO are
-included in the new "patent" directory.
-
-For details, please read the README.TXT in this directory.
-
-Distributions of the latest DDBJ release and newly-arrived/updated
-entries after that release can be retrieved at the following FTP
-sites.
-
-DDBJ Flat file
-ftp://ftp.ddbj.nig.ac.jp/ddbj_database/ddbj/
-ftp://ftp.ddbj.nig.ac.jp/ddbj_database/ddbjnew/
-
-DDBJ Flat file(TPA)
-ftp://ftp.ddbj.nig.ac.jp/ddbj_database/tpa/ddbj/
-ftp://ftp.ddbj.nig.ac.jp/ddbj_database/tpa/ddbjnew/
-
-INSD-XML
-ftp://ftp.ddbj.nig.ac.jp/ddbj_database/ddbj/xml/insdxml_current/
-ftp://ftp.ddbj.nig.ac.jp/ddbj_database/ddbjnew/xml/insdxml_current/
-
-It was last modified on Fri Feb 26 2010.
```

1 階層上のディレクトリへ

07/30/2010 06:07午後　ディレクトリ DRA000
07/05/2010 01:27午後　ディレクトリ ERA000
07/30/2010 08:09午後　ディレクトリ ERA005
07/30/2010 08:09午後　ディレクトリ ERA007
07/30/2010 08:09午後　ディレクトリ ERA008
07/30/2010 08:09午後　ディレクトリ ERA009
07/30/2010 08:09午後　ディレクトリ ERA010
07/13/2010 04:54午前　ディレクトリ SRA000
07/13/2010 03:26午前　ディレクトリ SRA001

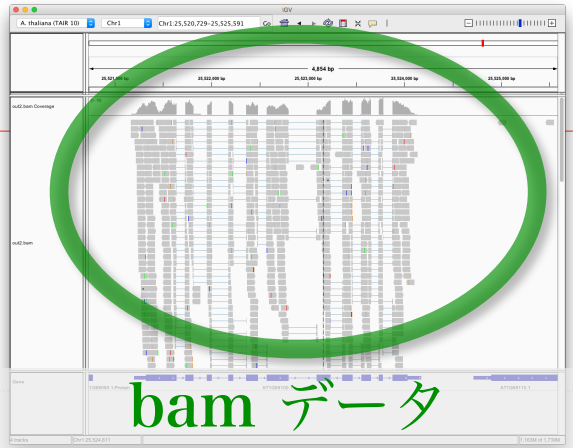
インターネット　100%

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

DRA の中身

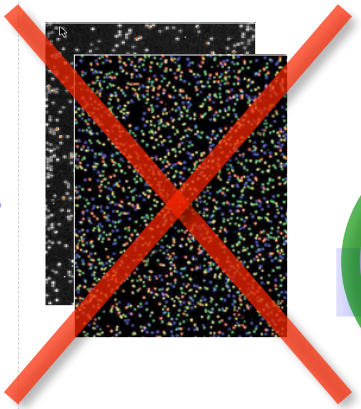
~~配列のみ~~



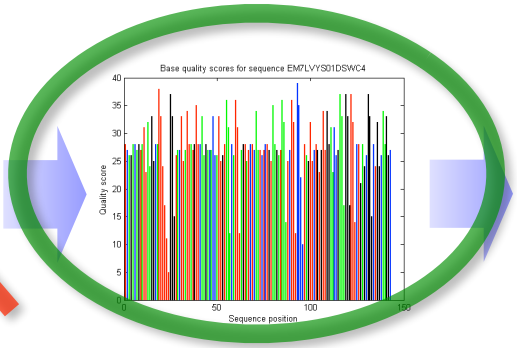
bam データ

アラインメントデータ

NGS



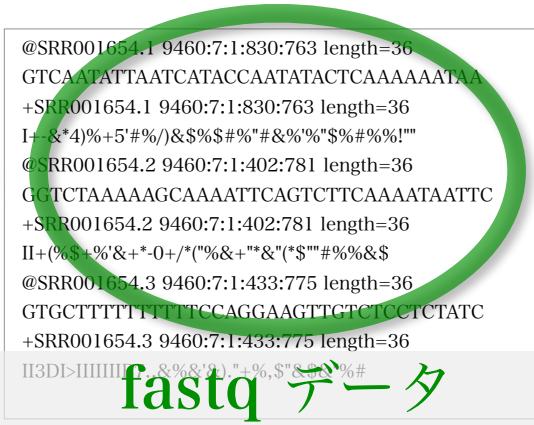
画像データ



1次データ

ベースコール

Quality etc



fastq データ

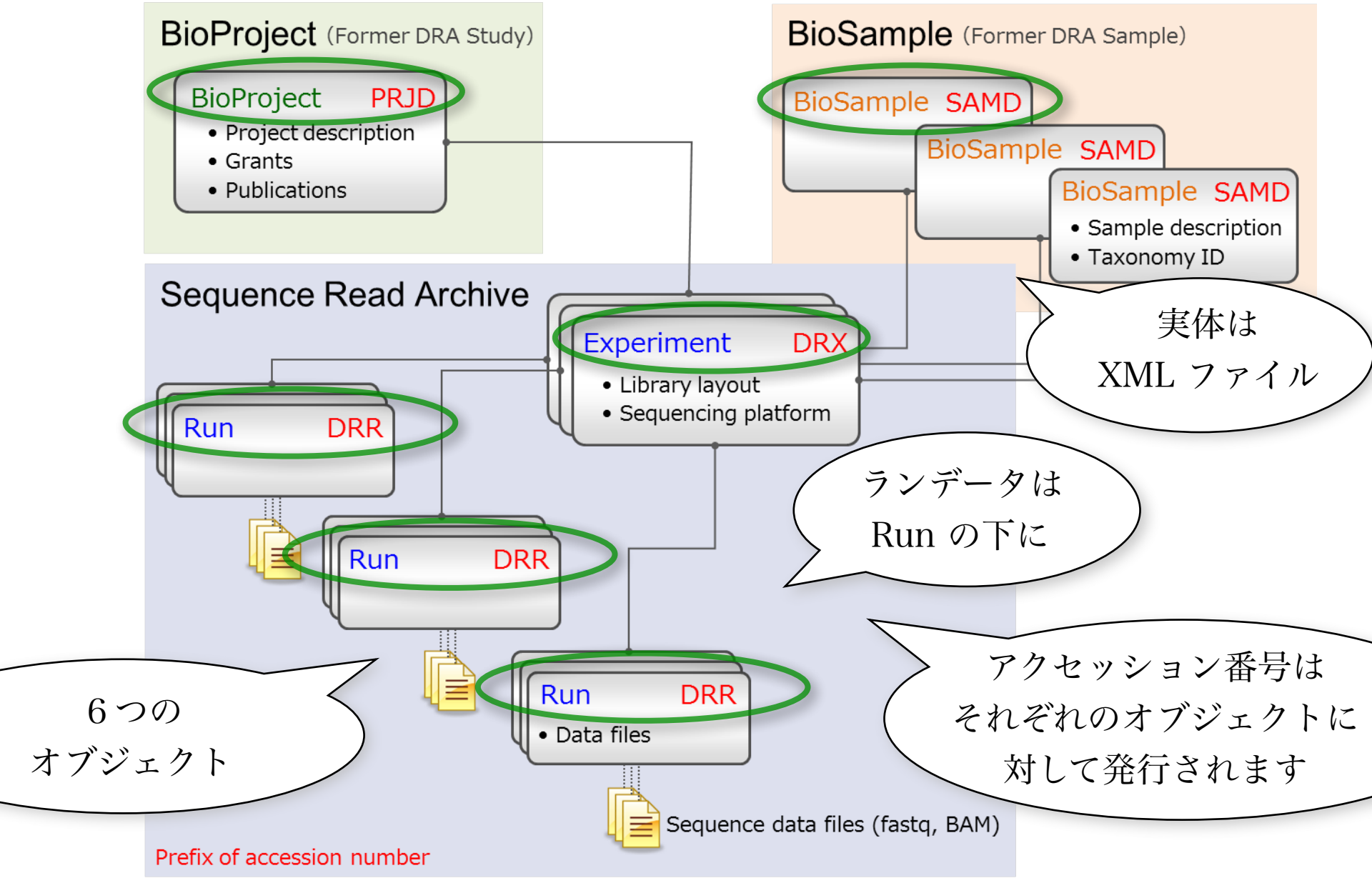
塩基配列 と

Quality Value

ランデータ

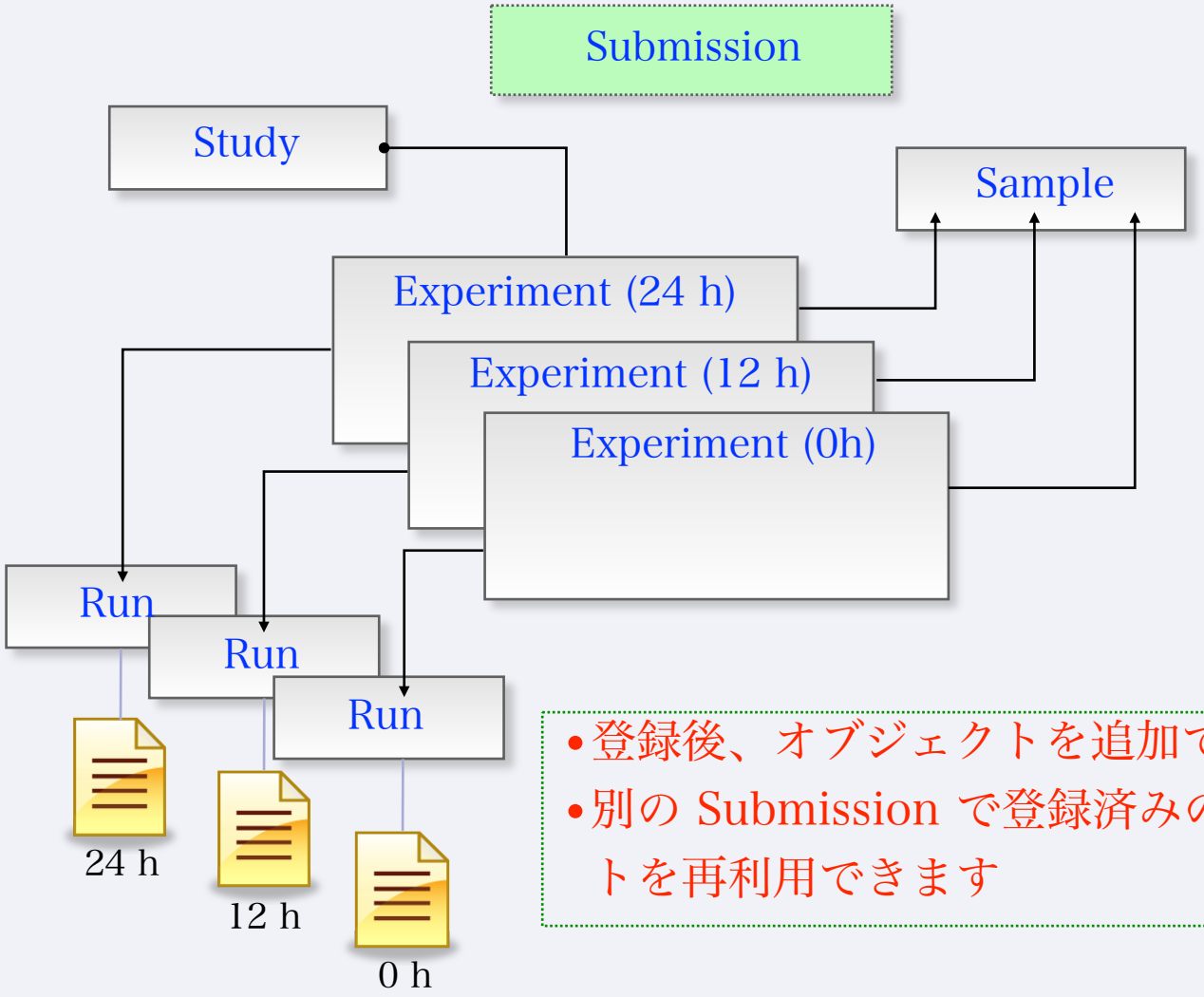
メタデータ (データに関する情報：実験手法、解析方法 etc)

メタデータとランデータの構成



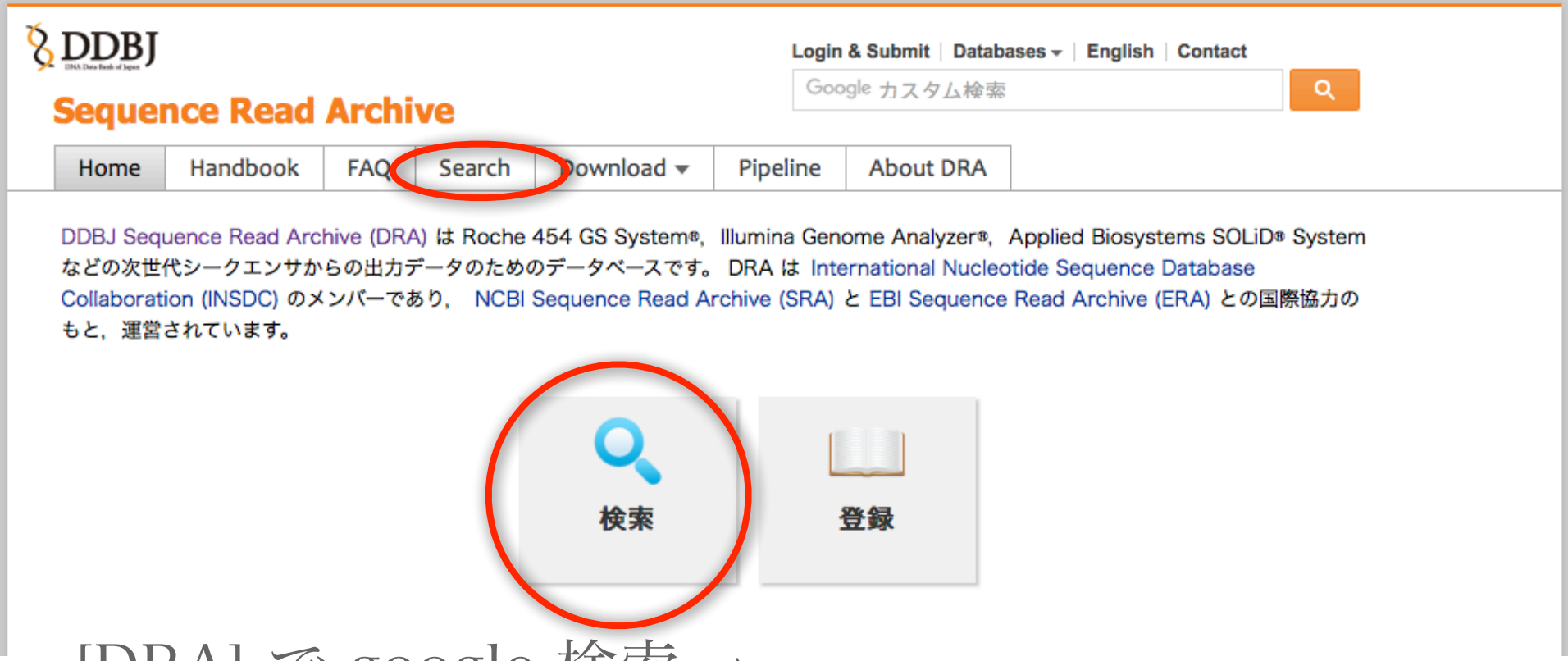
メタデータ構成の例

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

【実習】DRAsearch で検索してみよう



[DRA] で google 検索 ⇒

DDBJ Sequence Read Archive のサイトを開く

DRAsearch: *A. thaliana* を選択

DDBJ DNA Data Bank of Japan

DRASearch Search Home | DRA Home

Accession :

Organism : Arabidopsis t — StudyType : Transcriptome Analysis

CenterName : rm :

- Arabidopsis thaliana
- Arabidopsis thaliana x Arabidopsis arenosa
- Arabidopsis thaliana x Arabidopsis halleri
- Arabidopsis thaliana x Arabidopsis lyrata

Keyword :

Show 20 re ear

Statistics

Data Last Update 2017-08-23

Released Entries

Type	Count
Submission	731969
Study	112803
Experiment	3187852
Sample	2855310
Run	3593768

Organism

#	Organism Name	Study
1	Homo sapiens	10517
2	Mus musculus	8290
3	soil metagenome	2886
4	Oryza sativa	1341
5	Drosophila melanogaster	1333
6	Arabidopsis thaliana	1312
7	marine metagenome	1294
8	Populus trichocarpa	1088
9	Neurospora crassa	984
10	Saccharomyces cerevisiae	950

Study Type

#	Study Type	Study
1	Whole Genome Sequencing	45282
2	Other	36478
3	Transcriptome Analysis	15382
4	Metagenomics	13570
5	Population Genomics	787
6	Epigenetics	703
7	Exome Sequencing	231
8	Transcriptome Sequencing	169
9	Cancer Genomics	115
10	Pooled Clone Sequencing	35

Center Name

#	Center Name	Study
1	BioProject	61272
2	GEO	18494
3	UMIGS	2557
4	DOE - JOINT GENOME INSTITUTE	2485
5	JGI	2365
6	WUGSC	1394
7	JCVI	1148
8	BI	962
9	SC	904
10	The Wellcome Trust Sanger Institute	689

Website policy | © DNA Data Bank of Japan

Last modified: Oct. 13, 2016 (V3.1)

DRAsearch: Transcriptome を選択

DDBJ DNA Data Bank of Japan

DRASearch Search Home | DRA Home

Accession :

Organism : Arabidopsis thaliana

CenterName :

StudyType :

Platform :

Keyword :

Show 20 records Sort by Study Search Clear

- ✓
- 16S pyrosequencing
- amplicon analysis
- Amplicon sequencing
- Cancer Genomics
- ChIP-seq
- CLIP-seq
- Epigenetics
- Exome Sequencing
- genome partial sequencing
- genome-wide SNP discovery
- MAP
- Metagenomics
- Metatranscriptomics
- multi-isolate
- Other
- pooled clone
- Pooled Clone Sequencing
- Population Genomics
- RAD-seq
- Random Shotgun Sequencing
- RRBS
- small RNA
- Synthetic Genomics
- Target sequencing
- Targeted sequencing
- Transcriptome Analysis
- Transcriptome Sequencing
- vector insertion site analysis
- Whole chromosome sequencing
- Whole genome bisulfite sequencing
- Whole Genome Sequencing
- Whole Genome Sequencing

Search Results (1312 studies)

<< < 1 / 66 Page > >>

#	STUDY	SUBMISSION	STUDY_TITLE	_TYPE	ORGANISM	BASES	SUBMITTED	CENTER_NAME
1	DRP000209	DRA000208	Arabidopsis Transcriptome	me	Arabidopsis thaliana	11.2G	2010-08-02	NIG
2	DRP000347	DRA000344	Identification of EMS-induced Causal Mutations i thaliana Accession by Whole Genome Sequencin	ome	Arabidopsis thaliana	11.6G	2011-02-18	NAIST_BS
3	DRP000422	DRA000414	Whole genome bisulfite sequencng analysis on t	me quencing	Arabidopsis thaliana Neurospora crassa OR74A	15.1G	2011-07-05	UT_SCI
4	DRP000428	DRA000420	Re-sequencing of A. thaliana ddm1 genome 18A	ome	Arabidopsis thaliana	19.2G	2011-07-13	NIG
5	DRP000429	DRA000421	Re-sequencing of A. thaliana ddm1 genome 18C	ome	Arabidopsis thaliana	19.4G	2011-07-13	NIG
6	DRP000430	DRA000422	Re-sequencing of A. thaliana ddm1 genome 18H	ome	Arabidopsis thaliana	21.2G	2011-07-13	NIG
7	DRP000431	DRA000423	Re-sequencing of A. thaliana ddm1 genome 18J	ome	Arabidopsis thaliana	17G	2011-07-13	NIG
8	DRP000432	DRA000424	Re-sequencing of A. thaliana ddm1 genome 18L	ome	Arabidopsis thaliana	17.6G	2011-07-13	NIG
9	DRP000899	DRA000865	FE, a phloem-specific Myb protein, regulates the mobile floral signal FT	ome	Arabidopsis thaliana	5.9G	2012-12-10	NAIST_BS
10	DRP001015	DRA000976	Mechanism for full-length RNA processing over i	me	Arabidopsis thaliana	8.1G	2013-04-19	OIST
11	DRP001104	DRA001060	Whole genome bisulfite sequencing of A. thalian empty vector.		Arabidopsis thaliana	41.5G	2013-06-27	NIG
12	DRP001105	DRA001061	Whole genome bisulfite sequencing of A. thaliana, FT-283, transformed with full length VANDAL21, Hiun.	Epigenetics	Arabidopsis thaliana	44.3G	2013-06-28	NIG
13	DRP001134	DRA001087	Whole genome sequencing of a backcrossed F2 plant of EMS-mutagenized Arabidopsis thaliana	Whole Genome Sequencing	Arabidopsis thaliana	4.5G	2013-07-18	IBRC
14	DRP001245	DRA001183	Small RNA sequences in transgenic Arabidopsis plants expressiong a viral RNA silencing suppressor of plantago asiatica mosaic virus.	small RNA	Arabidopsis thaliana	703.4M	2013-10-17	UT-GALS
15	DRP001756	DRA001683	mRNA-seq analysis for Arabidopsis max1-1 mutant	Transcriptome Sequencing	Arabidopsis thaliana	143.7G	2014-02-04	TUAGRI
16	DRP001761	DRA001688	RNA-seq of ros1dml2;3 triple mutant under ABA and osmotic stress treatment	Transcriptome	Arabidopsis	45.4G	2014-02-05	RS_SNU

DRAsearch: 検索結果リスト

DDBJ DNA Data Bank of Japan

DRASearch Search Home | DRA Home

Accession :

Organism : Arabidopsis thaliana StudyType : Transcriptome Analysis

CenterName : Platform :

Keyword :

Show 20 records Sort by Study Search Clear

Search Results (548 studies)

<< < 1 / 28 Page > >>

#	STUDY	SUBMISSION	STUDY_TITLE	STUDY_TYPE	ORGANISM	BASES	SUBMITTED	CENTER_NAME
1	DRP000209	DRA000208	Arabidopsis Transcriptome	Transcriptome Analysis	Arabidopsis thaliana	11.2G	2010-08-02	NIG
2	DRP001800	DRA001727	Transcriptomic analysis of Arabidopsos rfl2 mutant	Transcriptome Analysis	Arabidopsis thaliana	6.5G	2014-02-21	NUGSS
3	ERP001018	ERA070277	Study of genome-wide alternative polyadenylation in A. thaliana	Transcriptome Analysis	Arabidopsis thaliana	326.9M		DUNDEE
4	ERP001616	ERA146059	CPL phosphatases as plant miRNA-processing regulators - mRNA reads	Transcriptome Analysis	Arabidopsis thaliana	18.2G		MPI-TUEBINGEN
5	ERP001622	ERA146132	CPL phosphatases as plant miRNA-processing regulators - sRNA reads	Transcriptome Analysis	Arabidopsis thaliana	7.7G		MPI-TUEBINGEN
6	ERP002233	ERA194802	Single-cell RNA-seq for pGL2 and QC cells from the Arabidopsis thaliana root	Transcriptome Analysis	Arabidopsis thaliana	89G		EMBL-EBI
7	ERP002617	ERA212631	Salt priming A. thaliana RNA-Seq	Transcriptome Analysis	Arabidopsis thaliana	10G		University of Glasgow
8	ERP004034	ERA256782	arabidopsis_thaliana_floral_meristem_atlas	Transcriptome Analysis	Arabidopsis thaliana	15.8G		University of Milan. Dept. of Biosciences
9	ERP005391	ERA295485	Transcriptome profiling of Arabidopsis embryogenesis in vitro	Transcriptome Analysis	Arabidopsis thaliana	21.3G		SCHOOL OF AGRICULTURE, POLICY AND DEVELOPMENT, THE UNIVERSITY OF READING, UK
10	ERP005565	ERA300733	Transcriptomic profiling of Arabidopsis DNA POLYMERASE EPSILON CATALYTIC SUBUNIT A (AtPOL2a) and B (AtPOL2b) mutants	Transcriptome Analysis	Arabidopsis thaliana	45.4G		SCHOOL OF AGRICULTURE, POLICY AND DEVELOPMENT, THE UNIVERSITY OF READING, UK
11	ERP006258	ERA326935	Arabidopsis thaliana RNase H2 deficiency counteracts the needs for the WEE1 checkpoint kinase but triggers genome instability	Transcriptome Analysis	Arabidopsis thaliana	11.7G		European Bioinformatics Institute
12	ERP006353	ERA330074	Extensive small RNAs sequestration in diverse eukaryotes	Transcriptome Analysis	Arabidopsis thaliana Caenorhabditis elegans Drosophila melanogaster Mus musculus Nicotiana	11.9G		EBI

DRAsearch: フリーキーワードで検索

DDBJ DNA Data Bank of Japan

DRASearch Search Home | DRA Home

Accession :

Organism : StudyType :

CenterName : Platform :

Keyword : Arabidopsis thaliana altanative splicing

Show 20 records Sort by Study Search Clear

Search Results (6367 records)

<< < 1 / 319 Page > >>

Filtered by

document type:study(2812) sample(1858) experiment(1393) submission(177) run(122) analysis(5)

organism:Arabidopsis thaliana(4150) Homo sapiens(460) Mus musculus(289) rhizosphere metagenome(151) Brassica rapa(110) Drosophila melanogaster(70)

#	META_FILE	ACCESSION	STUDY	STUDY_TITLE	STUDY_TYPE	ORGANISM	BASES	SUBMITTED	CENTER_NAME
1	SRA050132.study.xml ***splicing*** in ***Arabidopsis thaliana*** defense against the bacterial pathogen Pseudomonas syringae pv tomato (Pst	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	SRA200829.submission.xml in ***Arabidopsis thaliana***" lab_name="Bioinformatics core" center_name="IPMB" accession="SRA200829	SRA200829	SRP049707	Arabidopsis thaliana Transcriptome or Gene expression	Transcriptome Analysis	Arabidopsis thaliana	67.1G		BioProject
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	SRA198305.study.xml ***splicing***-contributed proteome diversity in ***Arabidopsis thaliana***</STUDY_TITLE> <STUDY_TYPE existing	SRP049534	SRP049534	Transcriptome survey of alternative splicing-contributed proteome diversity in Arabidopsis thaliana	Transcriptome Analysis	Arabidopsis thaliana	84.3G		BioProject
6	SRA200829.study.xml , that influences pre-mRNA ***splicing*** and is essential for embryonic development in ***Arabidopsis thaliana***. A genome	SRP049707	SRP049707	Arabidopsis thaliana Transcriptome or Gene expression	Transcriptome Analysis	Arabidopsis thaliana	67.1G		BioProject
7	SRA496773.study.xml </EXTERNAL_ID> </IDENTIFIERS> <DESCRIPTOR> <STUDY_TITLE>***Arabidopsis thaliana*** strain:Col-0	SRP093582	SRP093582	Arabidopsis thaliana strain:Col-0 Transcriptome or Gene expression	Transcriptome Analysis	Arabidopsis thaliana	0		BioProject
	SRA291694.study.xml			Arabidopsis thaliana strain:Columbia		Arabidopsis			

DRAsearch: Study Detail

DDBJ DNA Data Bank of Japan

DRASearch

Search Home | DRA Home

SRP000935

Study Detail	
Title	Transcriptome-wide map of alternative splicing in Arabidopsis
Study Type	Transcriptome Analysis
Abstract	placeholder
Description	Alternative splicing was profiled in Arabidopsis thaliana using Illumina-based RNA-seq. Five abiotic stress treatments and untreated controls were analyzed. All samples represented Col-0 wildtype plants.
Center Name	OSU-CGRB

Navigation			
Submission	SRA000931	FTP	
Experiment	SRX006191	FASTQ	SRA
	SRX006192	FASTQ	SRA
	SRX006193	FASTQ	SRA
	SRX006681	FASTQ	SRA
	SRX006682	FASTQ	SRA
	SRX006688	FASTQ	SRA
	SRX006690	FASTQ	SRA
	SRX006692	FASTQ	SRA
	SRX006704	FASTQ	SRA
	SRX179490	FASTQ	SRA
	SRX179491	FASTQ	SRA
	SRX179492	FASTQ	SRA
	SRX179493	FASTQ	SRA
	SRX179494	FASTQ	SRA
	SRX179495	FASTQ	SRA
Sample	SRS004095		
	SRS004097		
	SRS004098		
	SRS004099		
	SRS004100		
	SRS004101		
	SRS004102		
	SRS004103		
	SRS004104		

Website policy | © DNA Data Bank of Japan

DRAsearch: Experiment Detail

DDBJ DNA Data Bank of Japan

DRASearch

Search Home | DRA Home

SRX006191 FASTQ SRA

Experiment Detail	
Title	FC14
Design Description	Total RNA was extracted using "The Plant RNA Reagent" (Invitrogen). RNA was treated for 10 min at 65C with RNAsecure reagent (Ambion). To eliminate genomic DNA amplification, all RNA preparations were treated for 15 min at 37C with RNase-free "Turbo DNase" (Ambion). Next, total RNA was further purified using the "RNAeasy Mini RNA kit" (Qiagen) according to the manufacturer's RNA clean up protocol. Isolation of mRNA essentially free of ribosomal and other non-polyadenylated RNAs was critical for generation of non-biased randomly primed (RP) libraries. For the RP libraries the poly(A) mRNA was isolated by two consecutive cycles of purification on oligo d(T) cellulose using the "Micro-PolyA-Purist" kit (Ambion). Concentration, integrity and extent of contamination by ribosomal RNA were monitored using a ND-1000 spectrophotometer (Thermo Fisher Scientific) and Bioanalyzer 2100 (Agilent Technologies). Full-length enriched (FL) cDNA libraries were generated using the SMART method (Zhu et al., 2001). cDNA (5-7 ug in a 50 µL volume) was mixed with 750 uL of Illumina nebulization buffer and fragmented for 7 min in a nebulizer (Invitrogen) using compressed nitrogen at 35 psi. The sheared cDNA was purified using "QIAquick PCR Purification Kit" (Qiagen) and eluted into 32 uL of water. For the randomly primed (RP) cDNA libraries the first cDNA strand was synthesized using 1 ug of poly(A) mRNA essentially free of rRNA, random hexamer primers (300 ng per each ug of RNA), and Superscript III reverse transcriptase (RT) (Invitrogen). The second strand of cDNA was synthesized using DNA polymerase I (Klenow fragment) by combining 20 uL of the 1st strand reaction, 8 uL of 10x Klenow Buffer (NEB), 1 unit of RNAse H (Invitrogen), 68.8 uL of water and 30 units of DNA Polymerase I (NEB). The reaction was incubated for 90 min at 15C and cDNA was purified using "Qiaquick PCR clean up" kit. To prepare cDNAs for sequencing on the Illumina Genome Analyzer, FL or RP cDNA (30 uL) was combined with 10 uL of 10 mM ATP in 5X T4 DNA ligase buffer (Invitrogen), 4 uL of 10 mM dNTPs mix, 2.5 uL of T4 DNA polymerase (3 U/uL), 1 uL of Klenow DNA Pol (5 U/uL) and 2.5 uL of T4 polynucleotide kinase (10 U/uL, NEB). After incubation for 30 min at 20C the DNA was purified using "QIAquick PCR Purification kit". To add dA to the termini, 32 uL DNA from the prior step was mixed with 5 uL of 10X Klenow buffer, 10 uL of 1 mM dATP, and 3 ul of Klenow exo-polymerase (3' to 5' exo minus, 5 U/uL, NEB) and incubated for 30 min at 37C. DNA was purified using a "QIAquick MinElute Reaction Clean-up" (Qiagen) kit and eluted into 12 uL of water. To ligate Illumina adapters, 10 uL of cDNA from the prior step was mixed with 5 uL of 5X T4 DNA ligase buffer, 6 uL of adapter oligo mix, 4 uL of T4 DNA ligase (NEB) and incubated for 15 min at 25C. DNA was purified using a "QIAquick MinElute PCR Purification Kit". cDNA was size-fractionated (typically with average fragment length of 170 base pairs) on 3.5% (w/v) "NuSieve GTG" agarose. The fractionated libraries were PCR amplified using "Phusion Hot Start High-Fidelity" DNA polymerase (NEB) and the following PCR protocol: 30 sec at 98C, then 10 sec at 98C, 30 sec. at 65C, and 30 sec. at 72C for 18 cycles, followed by a 10-min extension step at 72C. Prior to cluster generation, DNA was diluted to a final concentration of 5 to 10 pM to generate 25,000 to 40,000 clusters per tile of the flow cell.
Organism	Arabidopsis thaliana

Navigation			
Submission	SRA009031	FTP	
Study	SRP000935		
Sample	SRS004095		
Run	SRR018346	FASTQ	SRA

例: この SRA データを使った二次論文

DDBJ DNA Data Bank of Japan

SRP000935 pubmed

ウェブ　画像検索　地図　ショッピング　もっと見

4 件 （0.19 秒）

Determinants of Exon-Level Evolutionary Rates in Ar

www.ncbi.nlm.nih.gov › ... › v.8; 2012 ▾ このページを訳す

GCT Wu 著 - 2012 - 引用元 1 - 関連記事

2012/07/04 - We retrieved a series of raw data submitted by Filic and Schimidt's lab (SRP007763). These RNA-seq data cover sev conditions, including abiotic stresses (cold, drought, heat, highlig

Estimation of Gene Expression at Isoform Level from

www.ncbi.nlm.nih.gov › ... › Front Genet › v.3; 2012 ▾ この

M Bhattacharjee 著 - 2012 - 関連記事

2012/11/27 - The dataset was downloaded from NCBI GEO webs **SRP000935**). Pooled data contains ..... [**PubMed**]; Bhattacharje Nelson P. S., Arjas E. (2004). Bayesian integrated functional ana

7075 - DNAnexus - SRA

sra.dnanexus.com/?page=7075&result_type... ▾ このペー

SRR019209, Oregon St Univ - Center for Genome Research & B SRX006688 · Arabidopsis thaliana · **SRP000935**, SRA009031. 1 Oregon St Univ ... **pubmed**(21393572)» · SRR099327, Stanford

Determinants of exon-level evolutionary rates in ... - F

Evol Bioinform Online. 2012; 8: 389–415.
PMCID: PMC3399485
Published online 2012 July 4. doi: 10.4137/EBO.S9743

Determinants of Exon-Level Evolutionary Rates in *Arabidopsis* Species

Gideon C.-T. Wu[1,2] and Feng-Chi Chen[1,2,3,4]

[1]Graduate Institute of Life Sciences, National Defense Medical Center, 114 Taiwan
[2]Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, 305 Taiwan
[3]Department of Life Sciences, National Chiao-Tung University, Hsinchu, 300 Taiwan
[4]Department of Dentistry, China Medical University, Taichung, 404 Taiwan
Corresponding author email: fcchen@nhri.org.tw

Front Genet. 2012; 3: 239.
PMCID: PMC3536024
Published online 2012 November 27. doi: 10.3389/fgene.2012.00239

Estimation of Gene Expression at Isoform Level from mRNA-Seq Data by Bayesian Hierarchical Modeling

M. Bhattacharjee,[1,2,*] Ravi Gupta,[3] and R. V. Davuluri[3]

[1]Department of Statistics, University of Pune, Pune, India
[2]Department of Mathematics and Statistics, University of Hyderabad, Hyderabad, India
[3]Center for Systems and Computational Biology, The Wistar Institute, Philadelphia, PA, USA
Edited by: Andrzej M. Kierzek, University of Surrey, UK
Reviewed by: Bartek Wilczynski, University of Warsaw, Poland; Gaurav Sablok, Istituto Agrario San Michele, Italy
*Correspondence: M. Bhattacharjee, Department of Mathematics and Statistics, School of Mathematics and CIS, University of Hyderabad, Prof. C.R.Rao Road, Hyderabad 500 046, Andhra Pradesh, India. e-mail:

お手軽解析パイプライン

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
DNA Data Bank of Japan

Select Query Files → Select Tools → Set QuerySet → Set GenomeSet → Set Ma

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 ANNOTA

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
☐	bwa		0.6.1	✓		✓	✓	✓	✓					✓	
☐	Bowtie		0.12.7	✓	✓	✓	✓	✓	✓	✓				✓	
☐	TopHat		1.0.11	✓		✓	✓	✓	✓					✓	
☐	Bowtie2		2.2.6	✓	✓	✓	✓	✓	✓	✓				✓	For reads longer than about 50 bp, Bowtie2 is generally faster, more senstitive, and uses less memory than Bowtie1.
☐	TopHat2		2.1.0	✓		✓	✓	✓	✓					✓	

de novo Assembly

Total limit = 22 Gbp

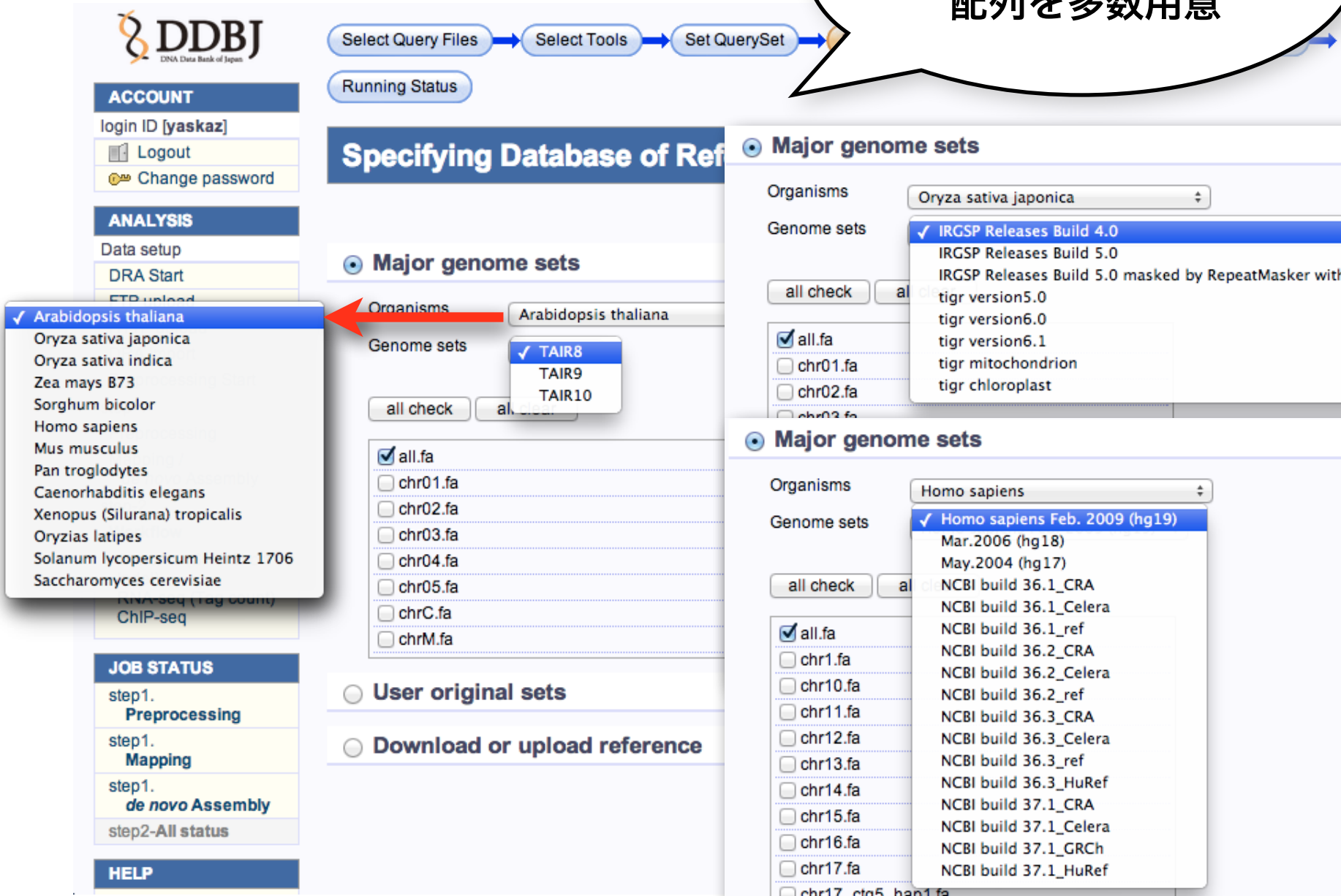
	Tool	Help	Version	Base space	Color space	Paired-end	MSS(WGS)	Comment
☐	SOAPdenovo		2.04-r240	✓		✓		
☐	ABySS		1.3.2	✓		✓		Maximum K-mer value is 64.
☐	Velvet		1.2.10	✓		✓	✓	We severe recommend when performing Velvet, total length of those reads is up to 22G bp.Maximum K-mer value is 64.
☐	Trinity		2.1.1	✓		✓		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)

良く使われるソフトウェアの特徴

マッピング	
BLAT	NGS以前からあるアライメントツール。発現データはイントロンを想定したギャップを考慮。
MAQ	NGS初期にショートリードに対応。リード長が長くなるに従い開発はBWAに引き継がれる。現在では DDBJ pipeline では提供していない。
BWA	MAQより速く、Titaniumのリードもオプションで対応。
SOAP	メモリ消費量少なく、高速。精度はBWAより弱冠落ちる。
Bowtie/Bowtie2	ギャップは考慮しないが処理は速い。BWA、SOAP2、Bowtieは Burrows-Wheeler変換というアルゴリズムでゲノムDNAにたいし てインデクスを作成、高速でマッピングする。Bowtie2は50bp以 上に最適化。
TopHat/TopHat2	RNA-Seqのリードを内部でBowtieを利用してマッピング、スプライスジャンクションを特定する。
アセンブル	
SOAPdenovo	ヒト、パンダ等大型ゲノムのアセンブリで使用された。比較的高速。
Abyss	初期に並列処理に対応したアセンブラ。
Velvet	NGS初期に開発された。メモリ消費多め。
Trinity	RNA-Seq配列用のアセンブラ。
Platanus	東工大で開発された国産アセンブラ。高速・メモリ消費少なめで、ヘテロなゲノムに強い。**おすすめ。**上記いずれも de bruijn graph というアルゴリズムを使用。
HGAP	PacBio データのための SMRT 法によるアセンブラ。

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



1. Mapping
2. Assembly

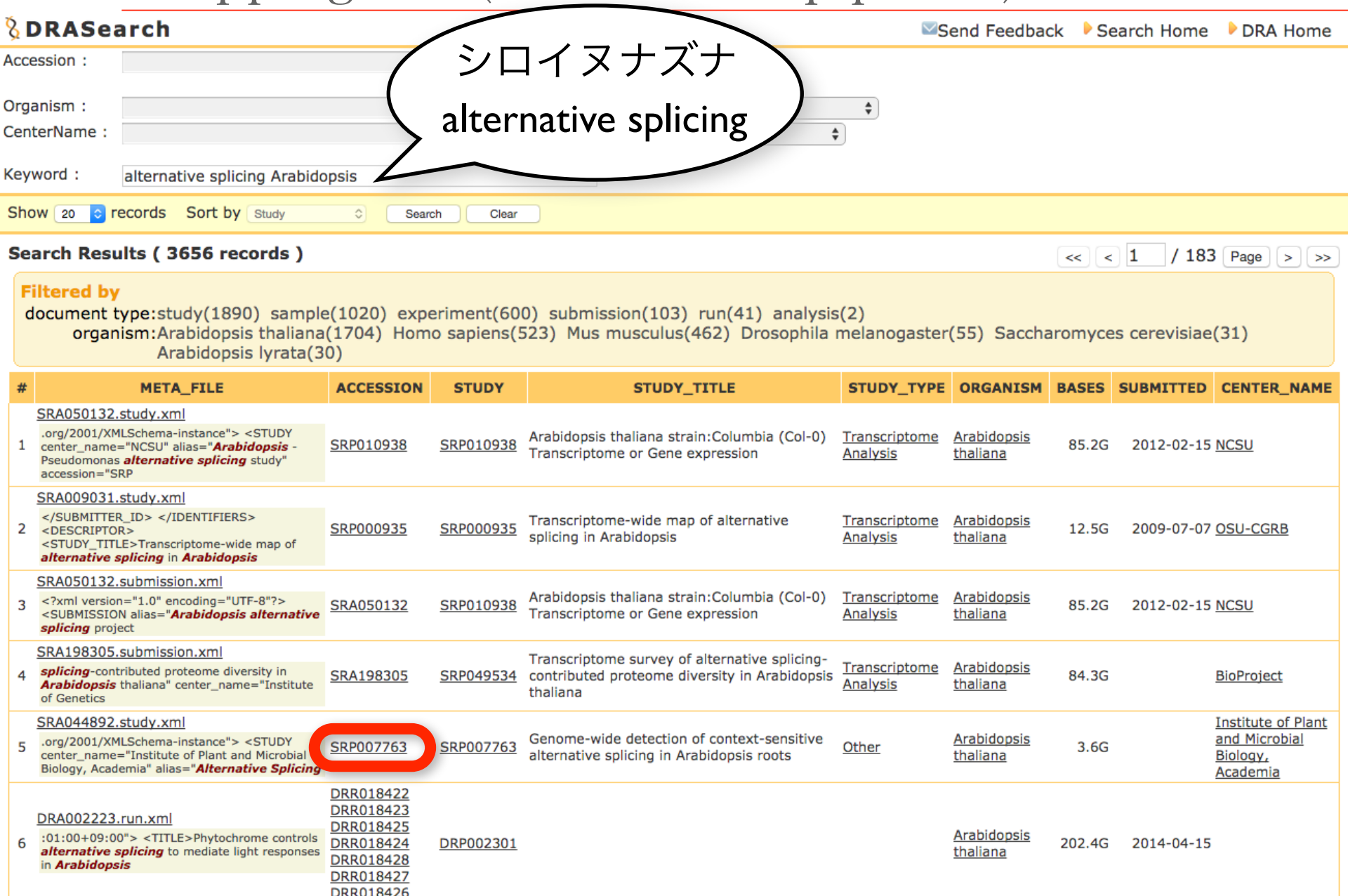
1. Mapping

2. Assembly

DRAウェブサイト ⇒ [DRA] で検索

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

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

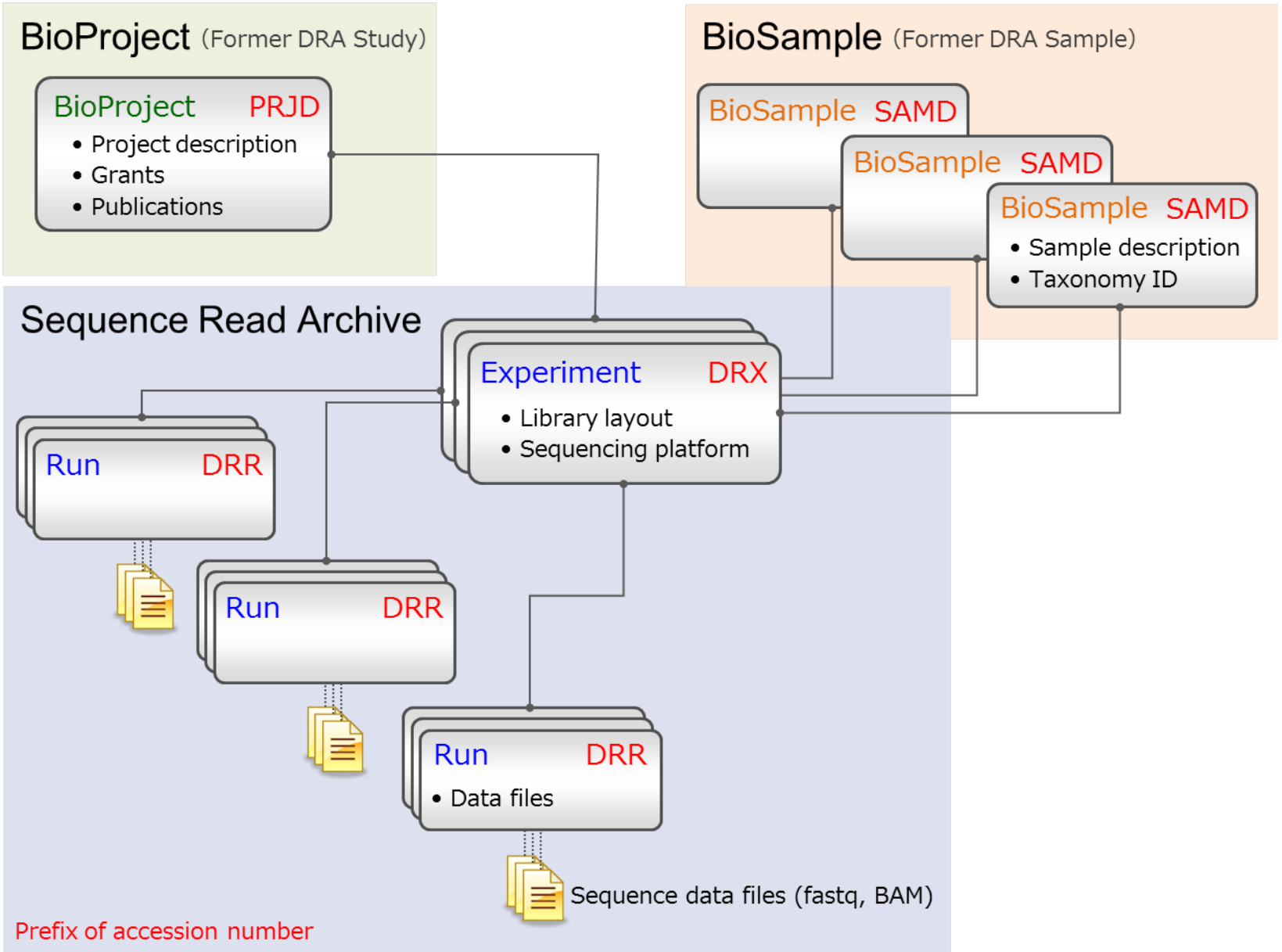
Study Detail	
Title	Genome-wide detection of context-sensitive alternative splicing in Arabidopsis roots
Study Type	Other
Abstract	To analyze context-sensitive changes in pre-mRNA splicing pattern and gene expression, we mapped the transcriptome of iron-deficient and iron-sufficient Arabidopsis roots 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		

Website policy | © DNA Data Bank of Japan

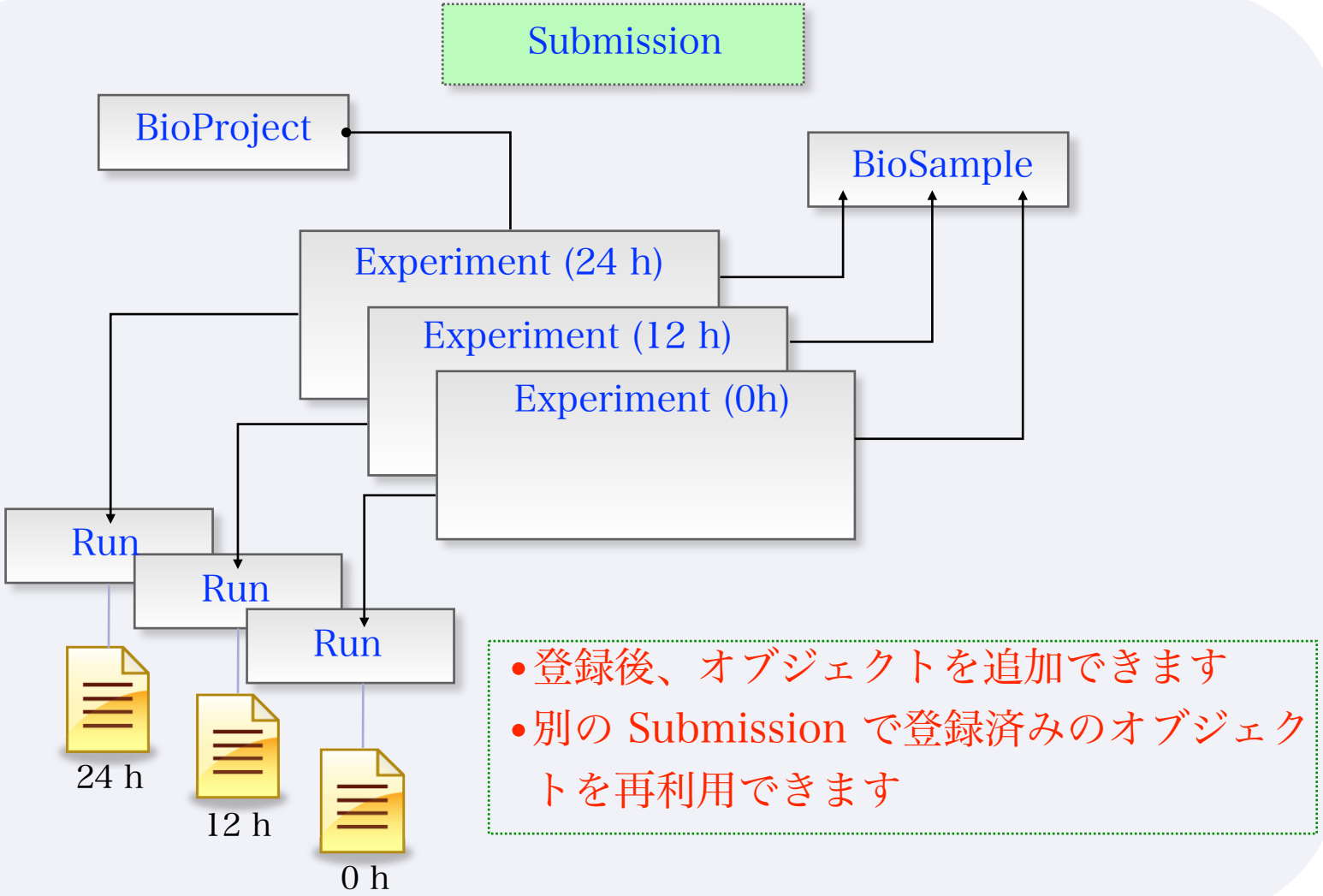
SRA の Metadata

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



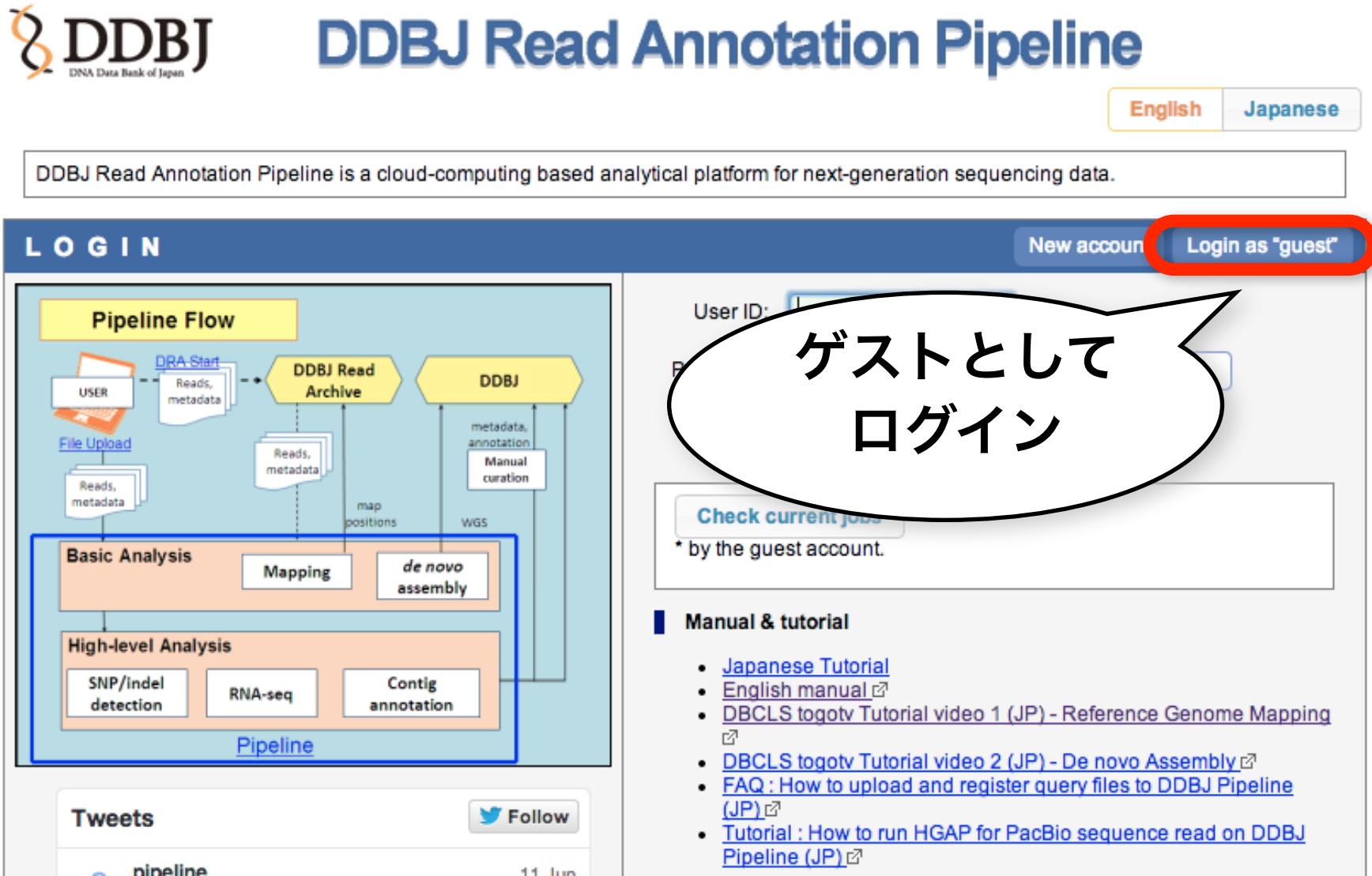
メタデータ構成の例

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

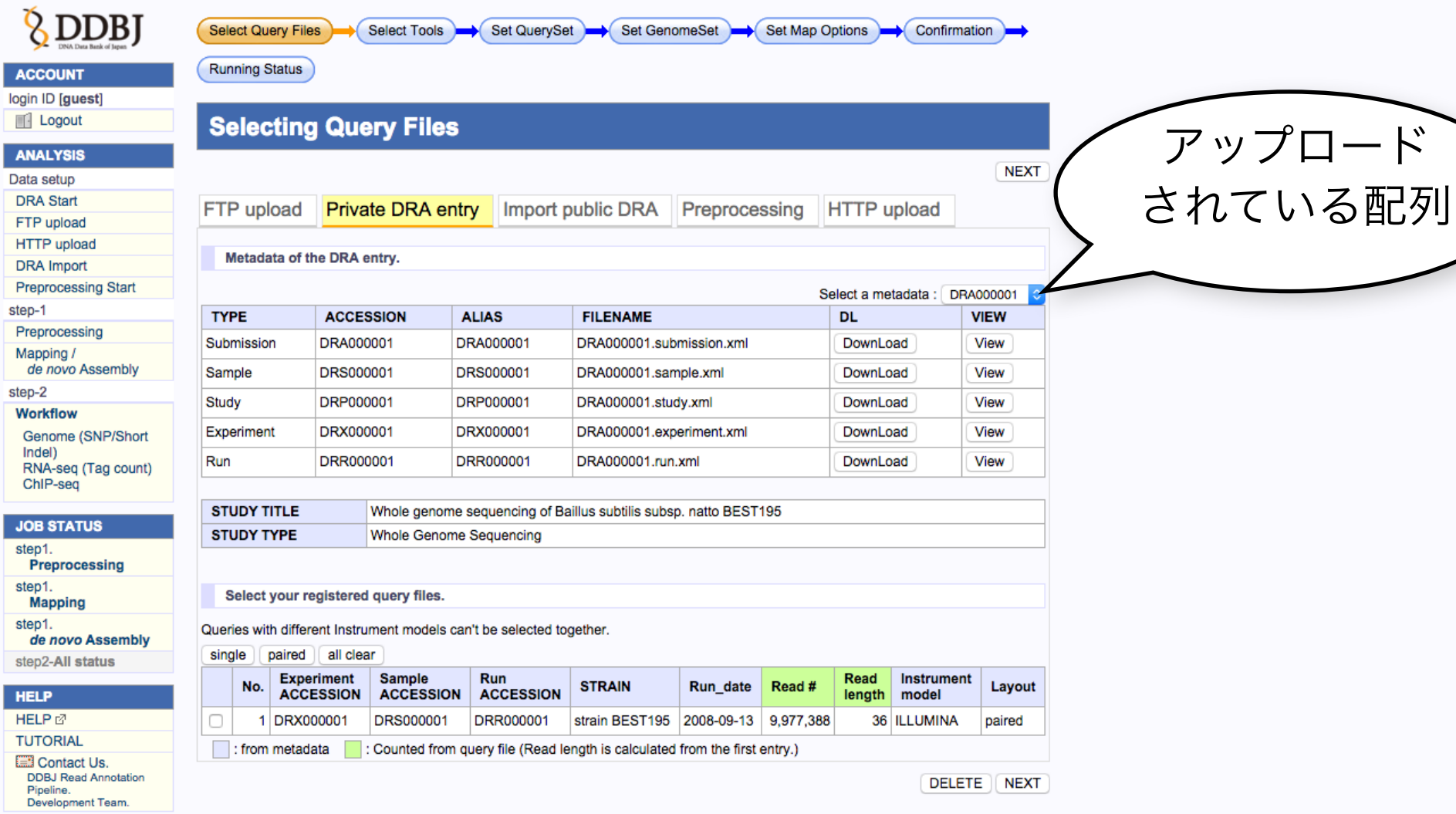


DDBJ パイプライン、体験してみましょう

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



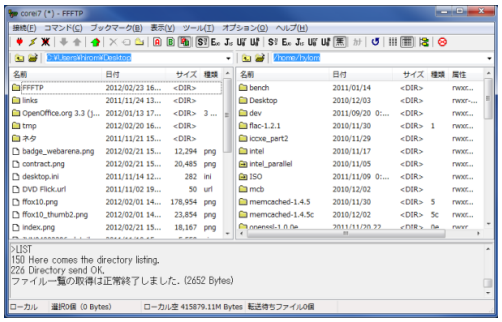
処理に使うNGSの配列ファイルの用意



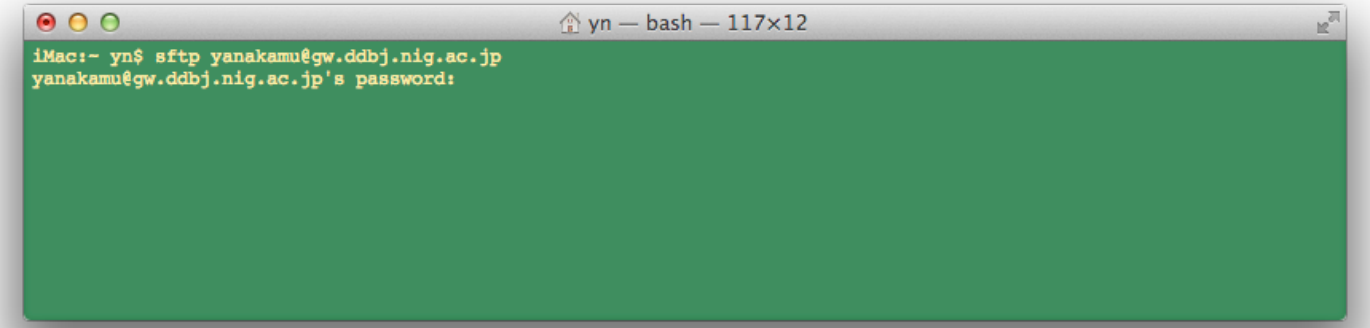
処理に使うNGSの配列ファイルの用意

FTP (file transfer protocol) クライアント

Cyberduck

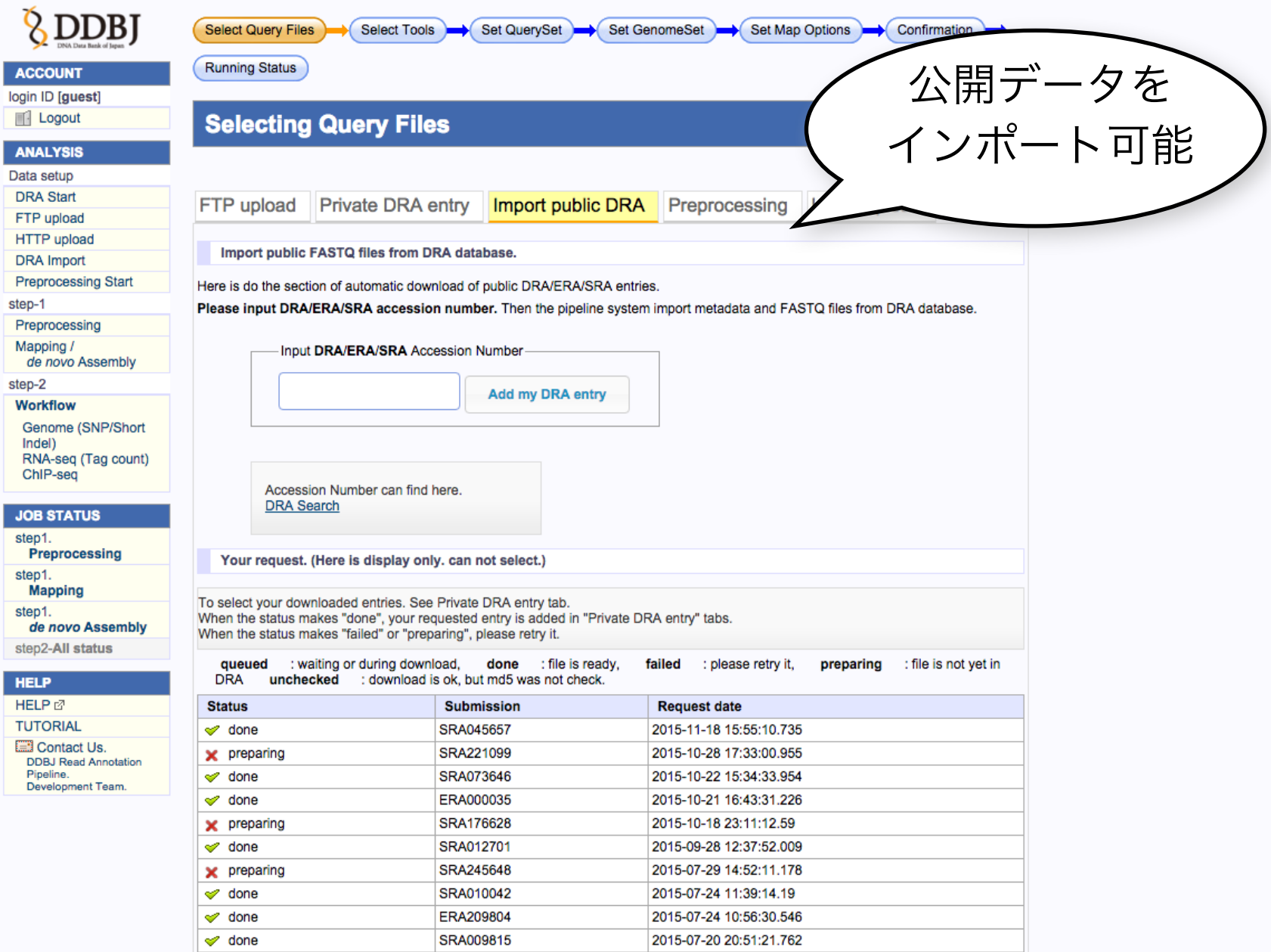


FFFTP

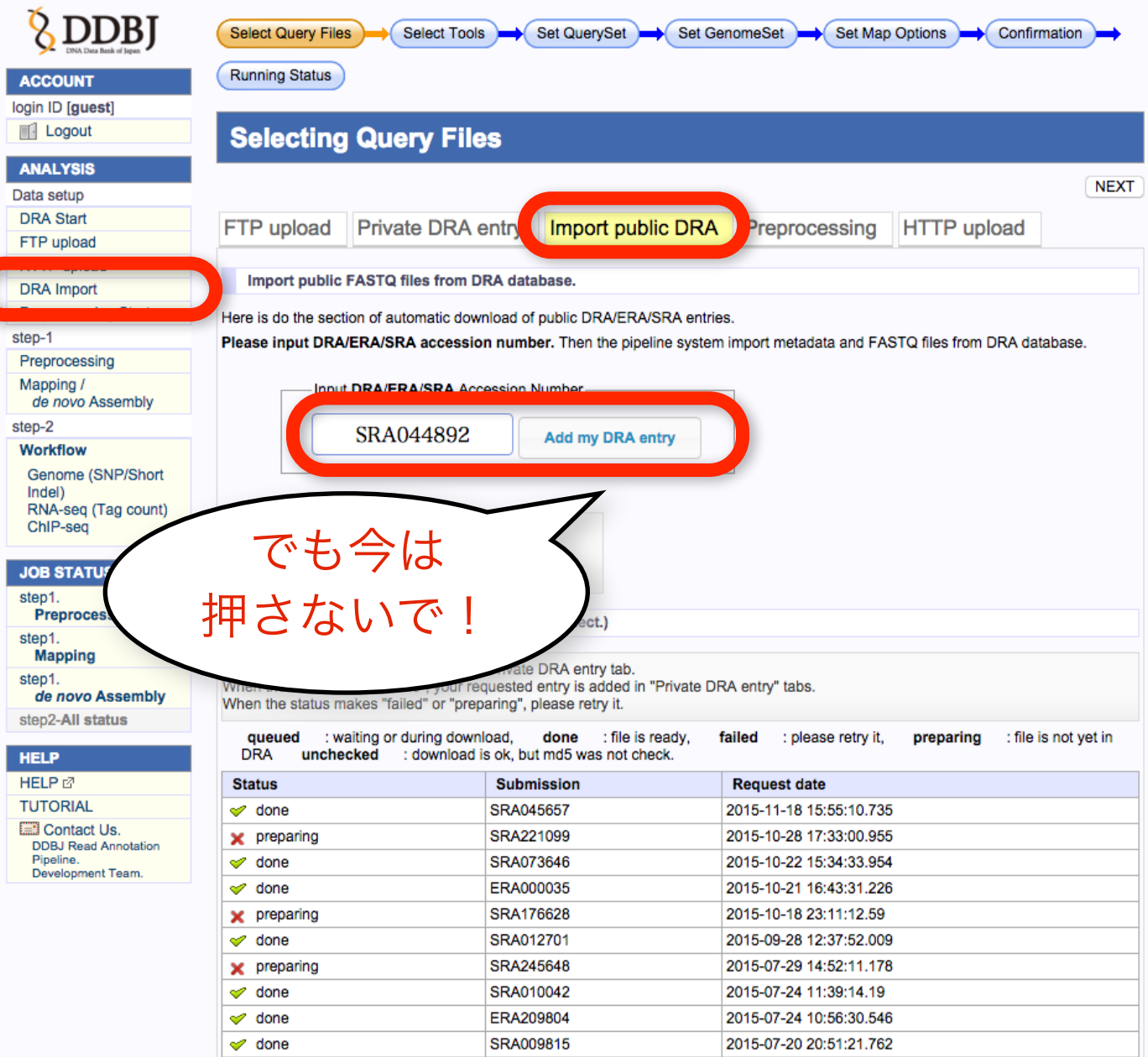


Cyberduck のような専用クライアントを使えばドラッグ＆ドロップでファイル転送可能。コマンドでの送受信もたいして難しくありませんが。

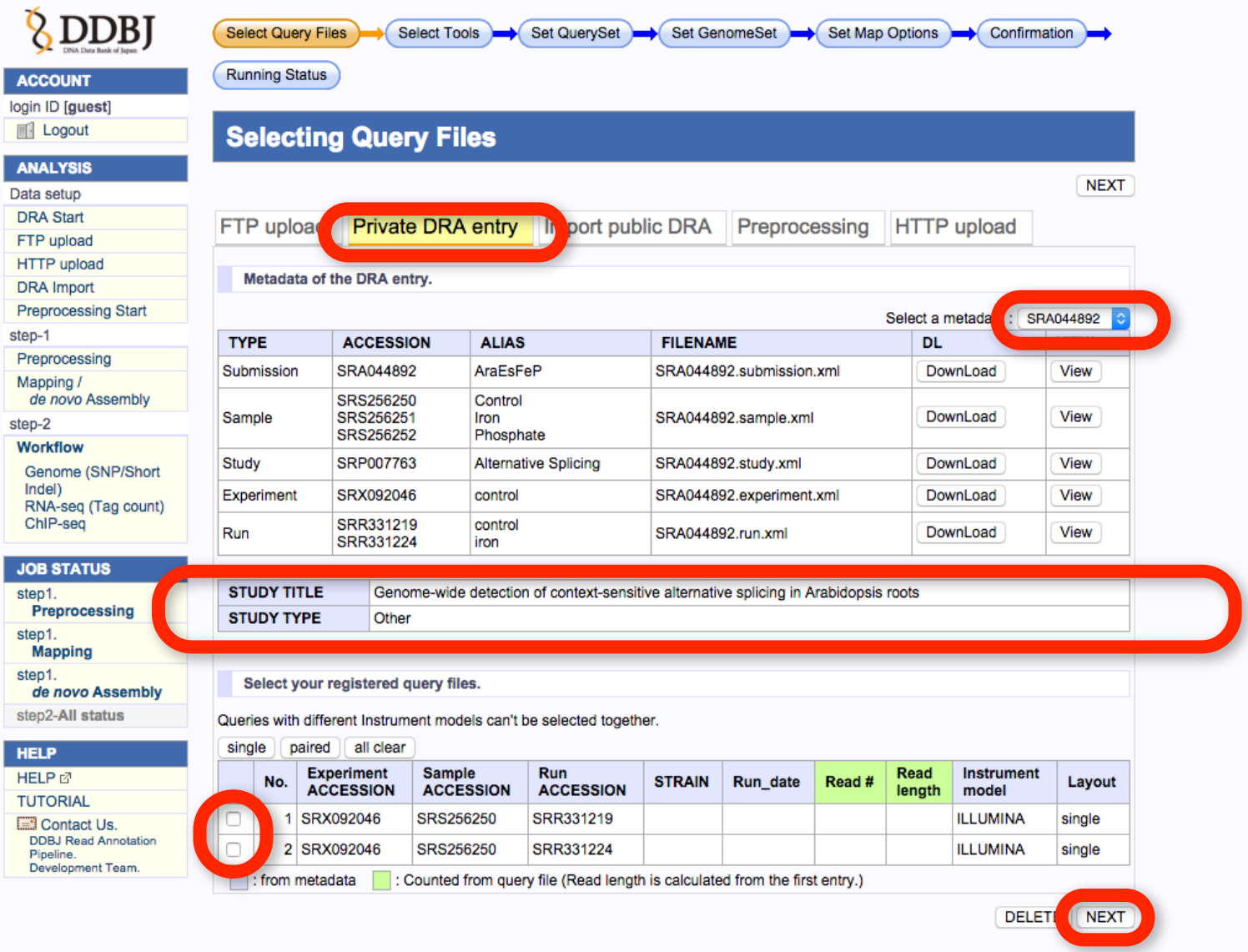
処理に使うNGSの配列ファイルの用意



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



Tophat2 を選んで NEXT

DDBJ
DNA Data Bank of Japan

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

				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
☐	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.
☐	[illegible]		2.0.9	✓		✓	✓	✓	✓					✓	

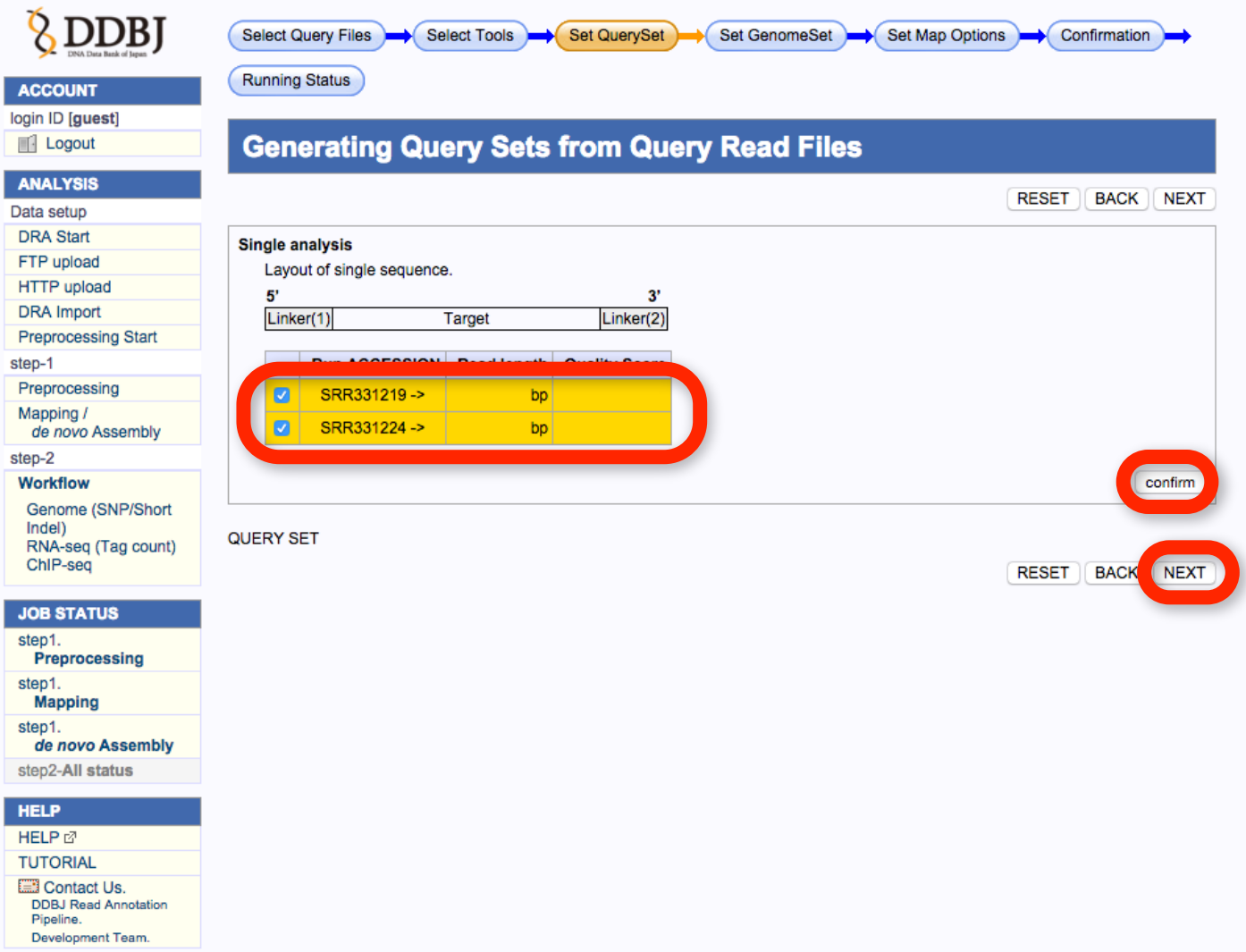
de novo Assembly

Total limit = 22 Gbp

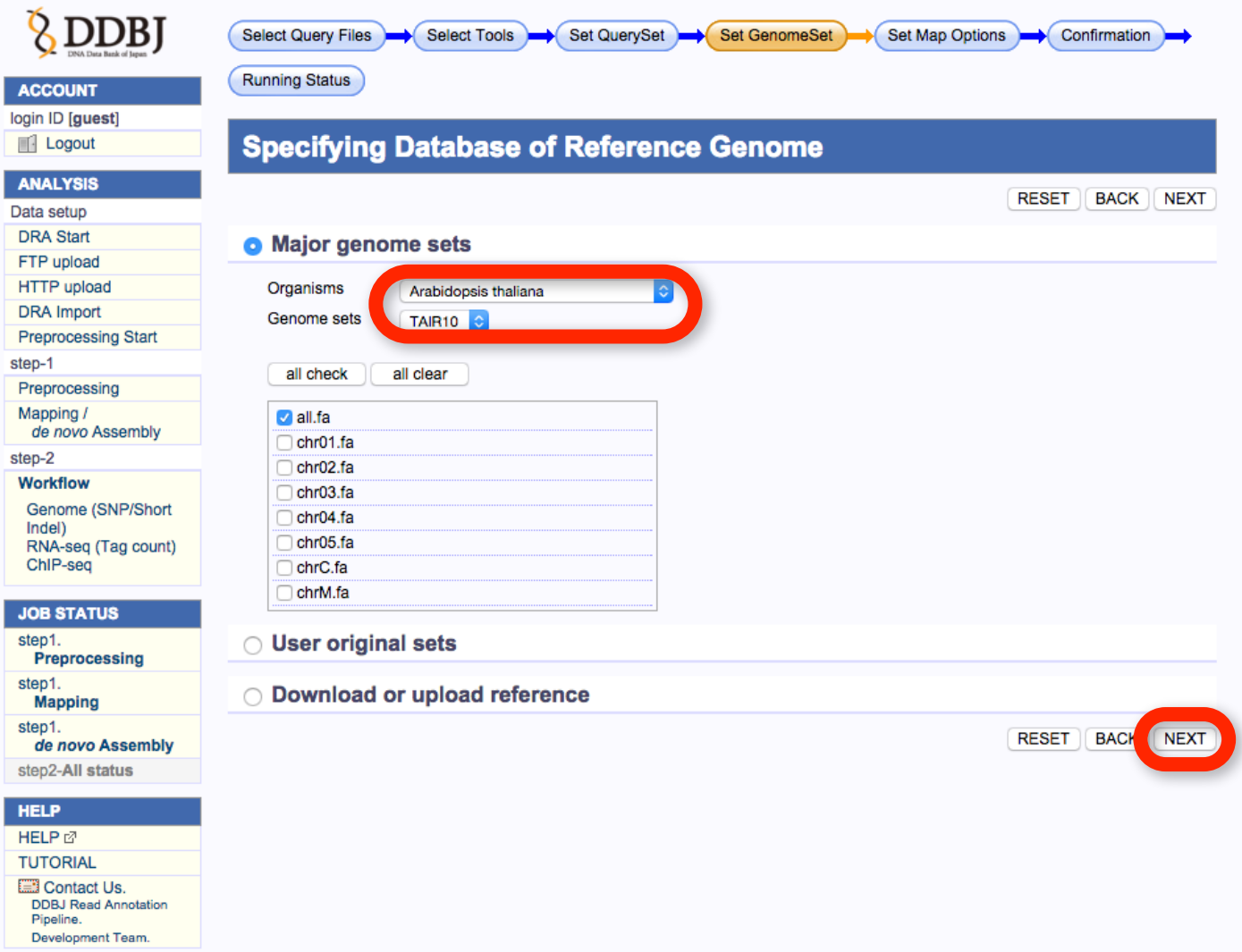
	Tool	Help	Version	Base space	Color space	Paired-end	MSS(WGS)	Comment
☐	SOAPdenovo		1.05	✓		✓		
☐	ABySS		1.3.2	✓		✓		Maximum K-mer value is 64.
☐	Velvet		1.2.10	✓		✓	✓	We severe recommend when performing Velvet, total length of those reads is up to 22G bp.Maximum K-mer value is 64.
☐	Trinity		r2013-02-25	✓		✓		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

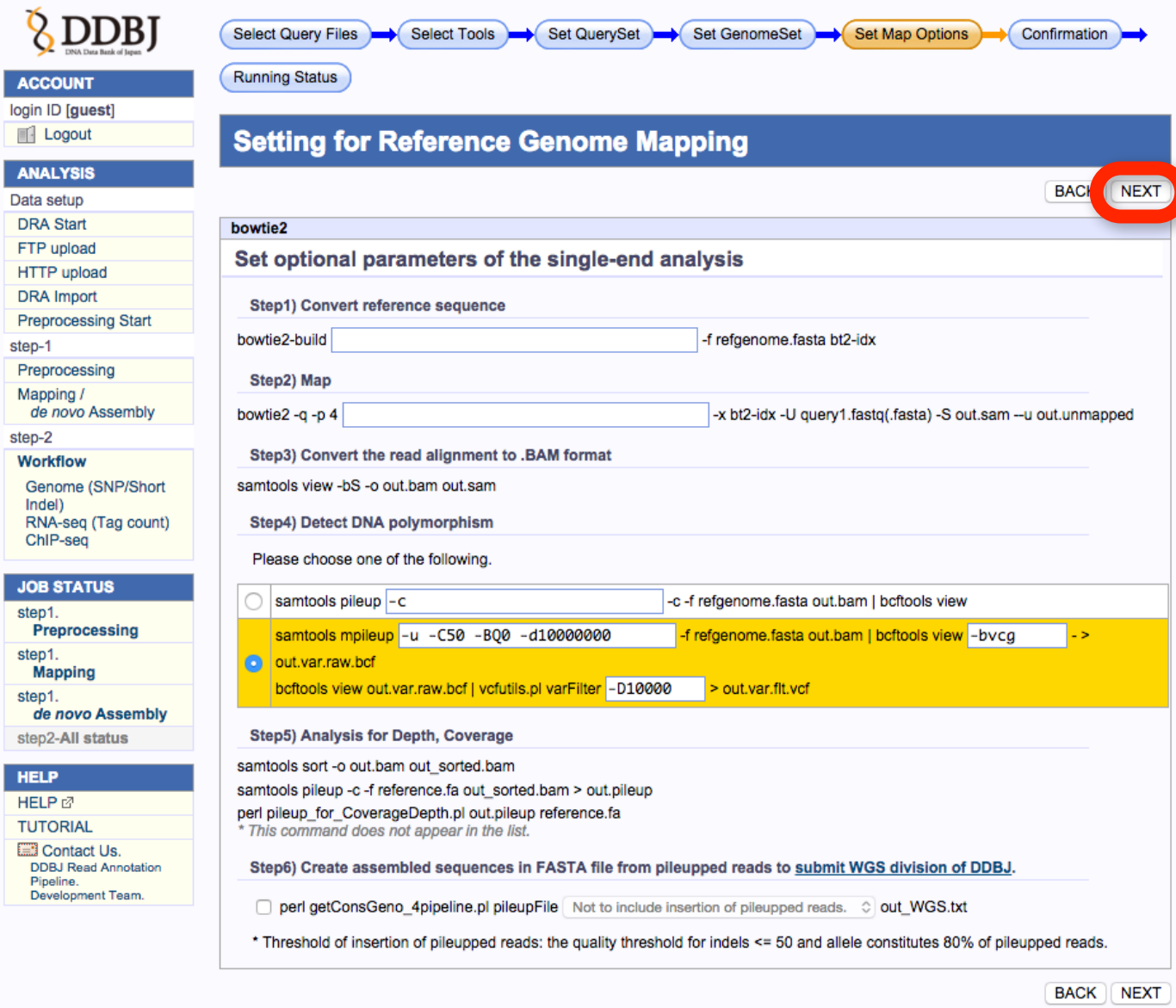
配列を選んで confirm, NEXT



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

DDBJ
DNA Data Bank of Japan

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.

Run Confirmation

BACK

Destination of mail

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

yn@nig.ac.jp * Required

Result files will be deleted 60 days after submission.

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

Step3) Convert the read alignment to .BAM format

samtools view -bS -o out.bam out.sam

Step4) Detect DNA polymorphism

Please choose one of the following.

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

samtools mpileup -u -C50 -BQ0 -d10000000 -f refgenome.fasta out.bam | bcftools view -bvc

out.var.raw.bcf

連絡先いれたら
RUN可能

guest は
run できません

pipeline_team@g.nig.ac.jp

Job finished : DDBJ Read Annotation Pipeline

宛先: 中村保一 <yn@nig.ac.jp>, CC: pipeline_report@g.nig.ac.jp

Message-Id:<599d04f7.4507630a.708a1.2bb9@mx.google.com>

Dear yaskaz,

Your request to DDBJ pipeline service has finished.
Please visit the web site to obtain analytical results.
Request ID: 27833
URL: https://p.ddbj.nig.ac.jp/

If you have troubles in this service, please write to

「RUN を押した」と思ってください

DDBJ
DNA Data Bank of Japan

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

out2.bam
out2.bam.bai

Detail view

BACK

Job info

ID
27833

Tool (Version)
TopHat2 (2.1.0)

RunAccession or Filename	Download	Read length	Alias
SRR331219	SRR331219.fastq.bz2	N.A. bp	control
SRR331224	SRR331224.fastq.bz2	N.A. bp	iron

Genome set
TAIR10

Chromosome
all.fa

Download modified queries

- mergedout.fastq.gz (Original size 4.6 GB)

Download merged pileup file

Uniq.sam files have been merged if you specified 'uniq' option.

- merged.var.flt.vcf.gz (Original size 923.6 KB)
- merged.sam.gz (Original size 4.3 GB)

Download wgs file

- out_WGS.fasta.gz (Original size 138.4 KB)

Position errors	Map ratio	Depth, Coverage
PDF download	total query # : 17,352,005 mapped query # : 15,596,984 map ratio : 89.886 %	coverage : 53774974 / 119482012 * 100 = 45.007 depth : 2452979062 / 53774974 = 45.616

Time

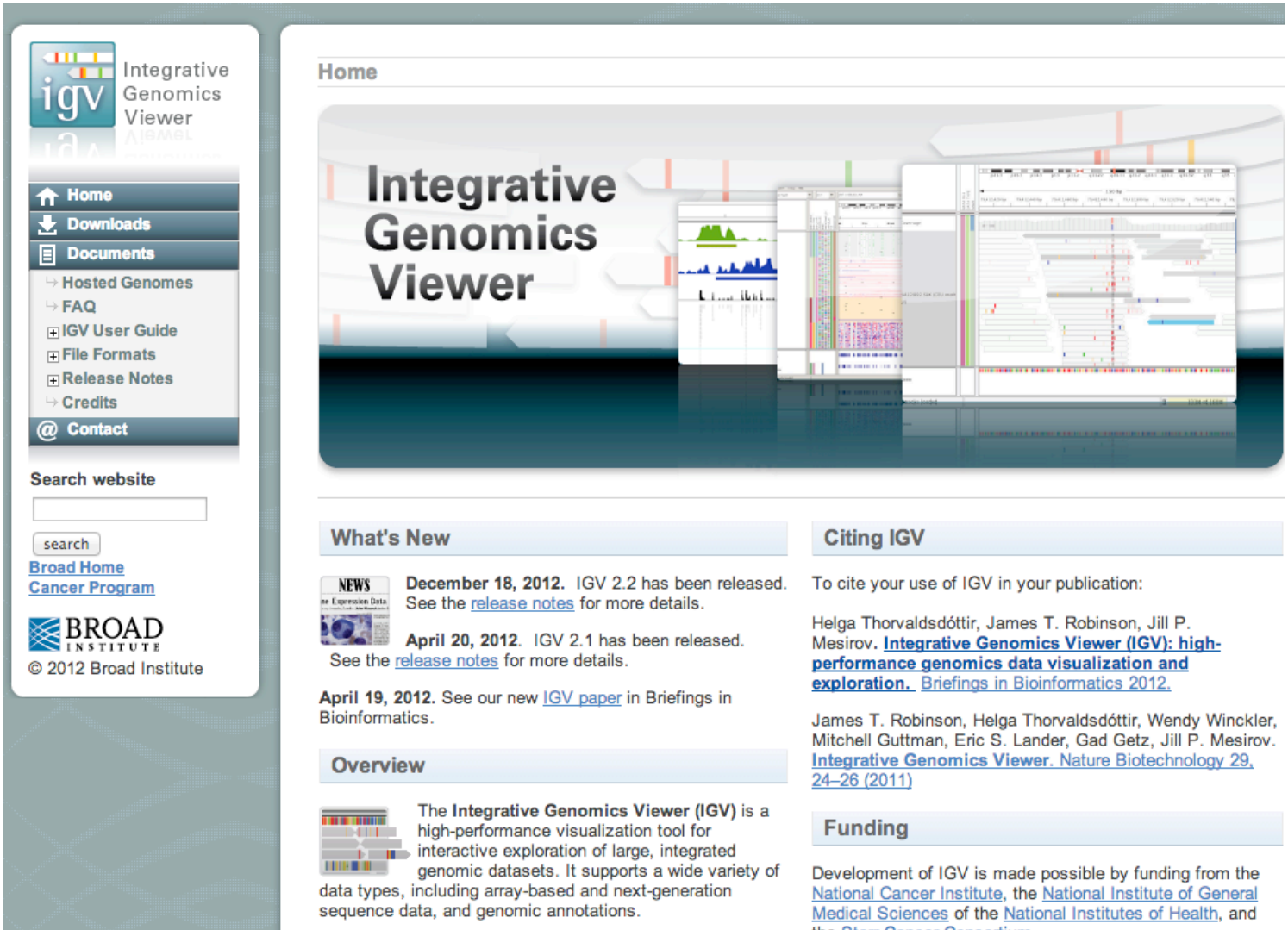
Wait time	Start time	End time
0: 3:35	2017-08-23 11:28:12	2017-08-23 13:30:44

all.fa

Command	Start time	End time	Log1	Log2	Result	MD5
Create Bowtie Index File bowtie2-build -f all.fa refgenome	2017-08-23 11:28:13	2017-08-23 11:30:58	View	View		
Map reads with TopHat tophat2 -p 4 -o result_directory refgenome mergedout.fastq	2017-08-23 11:30:58	2017-08-23 12:30:26		View	Download(992.0 MB)	MD5
Conver BAM to SAM samtools view -h accepted_hits.bam > out.sam	2017-08-23 12:31:18	2017-08-23 12:32:00				MD5
Sort BAM File samtools sort accepted_hits.bam out2	2017-08-23 12:35:16	2017-08-23 12:39:24			Download(891.3 MB)	
Create BAM Index File samtools index out2.bam	2017-08-23 12:40:05	2017-08-23 12:40:26			Download(110.4 KB)	
Mpileup and Create BCF File samtools mpileup -u -C50 -BQ0 -d10000000 -f all.fa out2.bam | bcftools view -bvcg - > non-uniq.var.bcf	2017-08-23 12:40:37	2017-08-23 12:59:42		View		
Filter BCF and Convert to VCF File bcftools view non-uniq.var.bcf | perl vcfutils.pl varFilter -D10000 > out.var.flt.vcf	2017-08-23 12:59:42	2017-08-23 12:59:53			Download(187.2 KB)	MD5
PileUp from Sorted BAM File For DepthCoverage samtools pileup -c -f all.fa out2.bam > out.pileup	2017-08-23 13:00:04	2017-08-23 13:14:50				
Sort BAM File For MapRatio samtools sort -n accepted_hits.bam out_sorted_by_name	2017-08-23 13:14:51	2017-08-23 13:21:54		View		
Convert BAM to SAMX For MapRatio samtools view -hX out_sorted_by_name.bam > out_sorted_by_name.samX	2017-08-23 13:21:54	2017-08-23 13:22:46				
Convert BAM to SAMX For ErrorRate samtools view -hX accepted_hits.bam > out.samX	2017-08-23 13:22:46	2017-08-23 13:23:27				

マッピング結果を眺めるビューワ

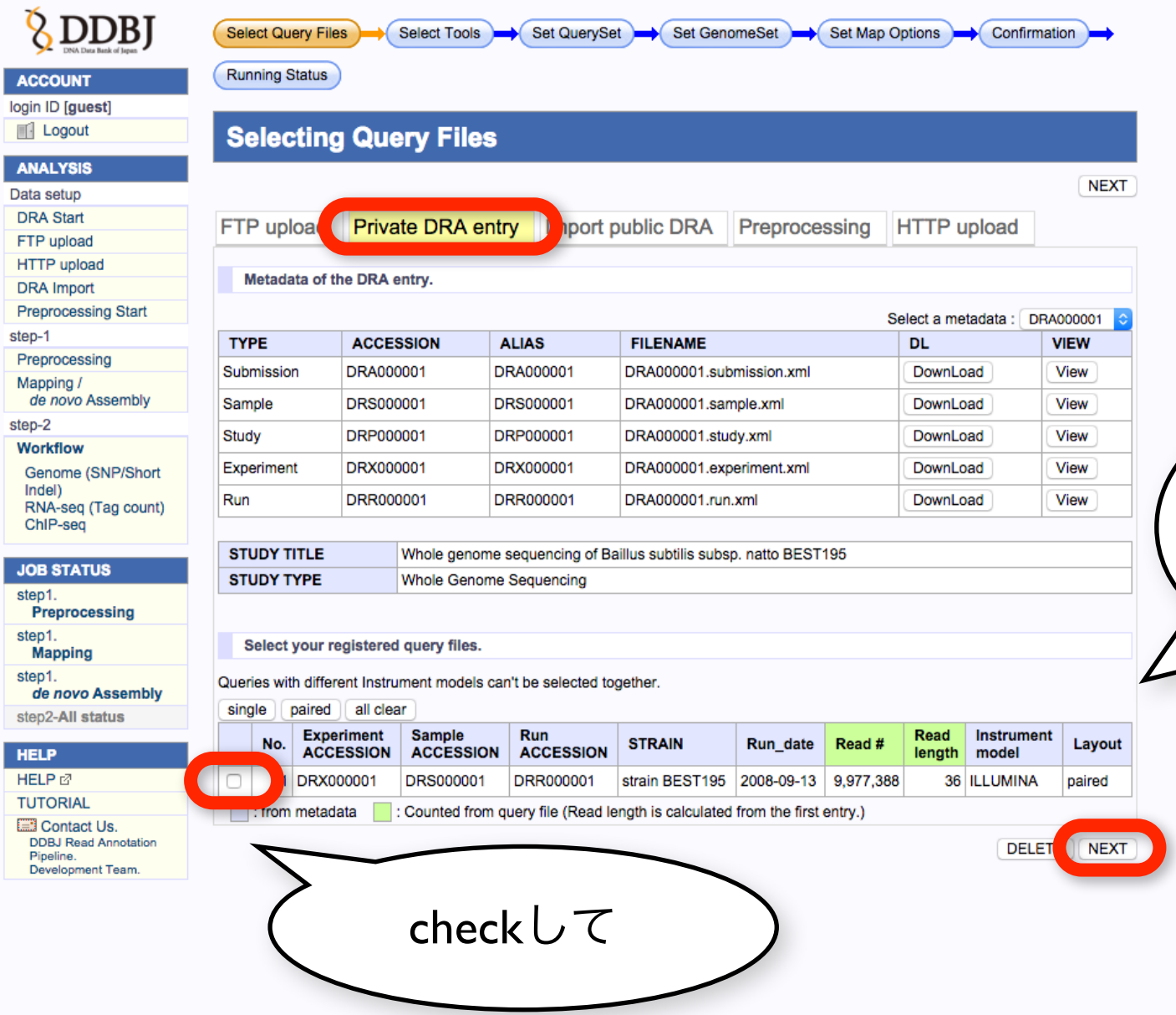
IGV (http://www.broadinstitute.org/igv/)



1. Mapping

2. Assembly

インポート済のエントリから



納豆菌の
公開データが
インポート済

checkして

DDBJ
DNA Data Bank of Japan

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.

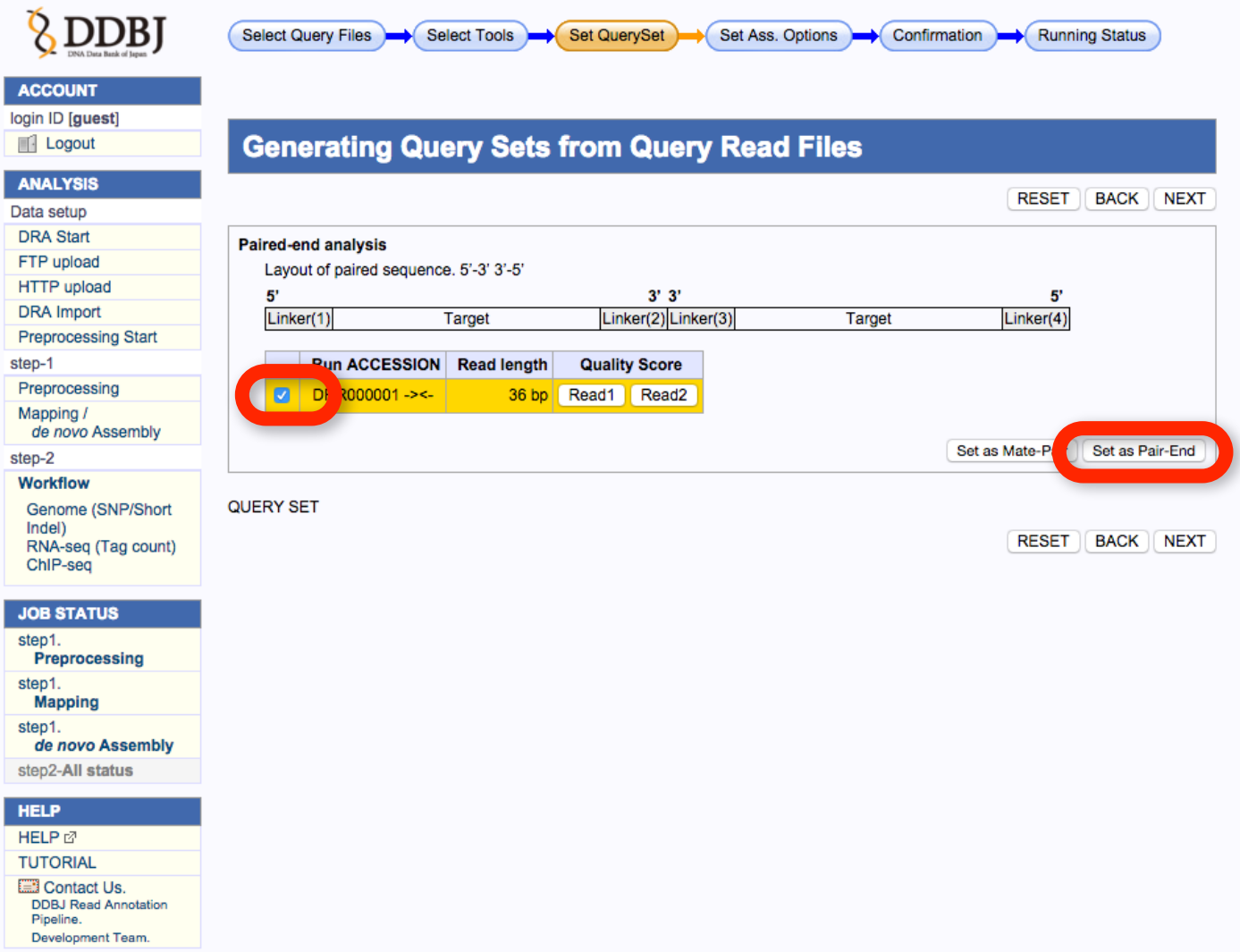
The contigs will be aligned to reference genome.

	Tool	Comment
◉	BLAT	Single-end analysis only

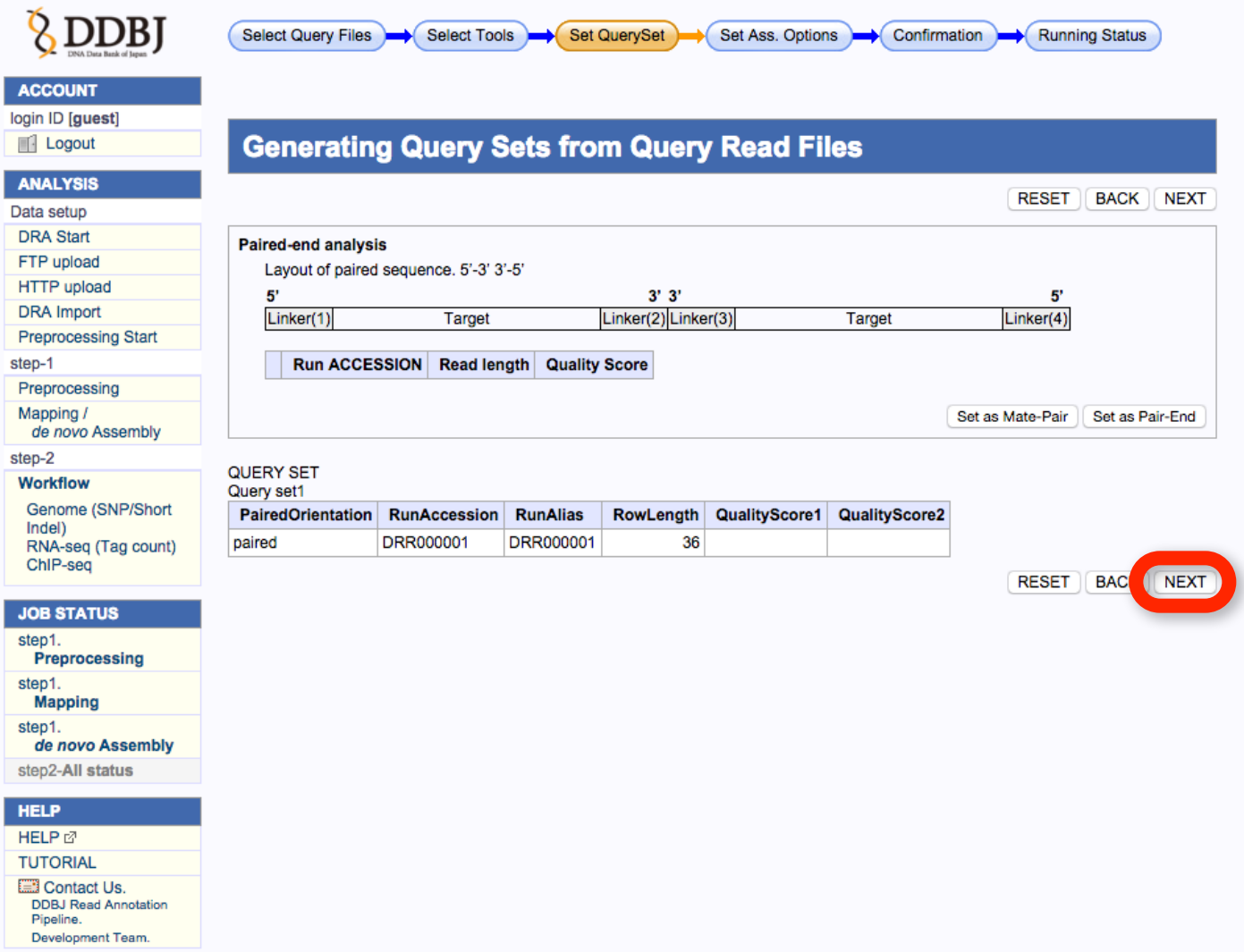
BAC NEXT

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

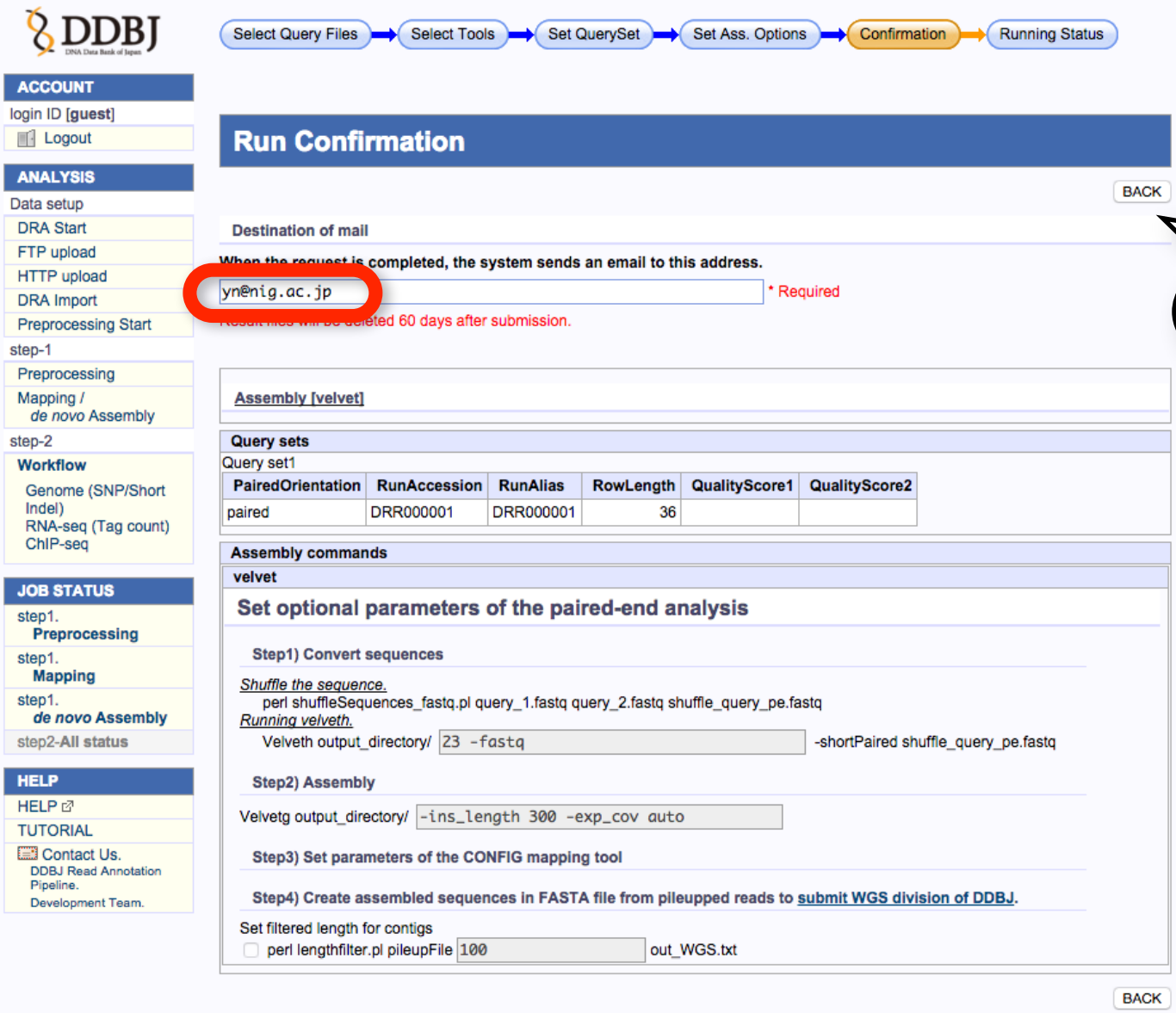


配列のセットの形式を選んで次へ



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

パイプライン利用申請はこちら

http://p.ddbj.nig.ac.jp　[DDBJ pipeline] で検索