

DDBJと新型シーケンサ 解析パイプラインの使い方

情報・システム研究機構
国立遺伝学研究所 DDBJセンター 教授
中村 保一

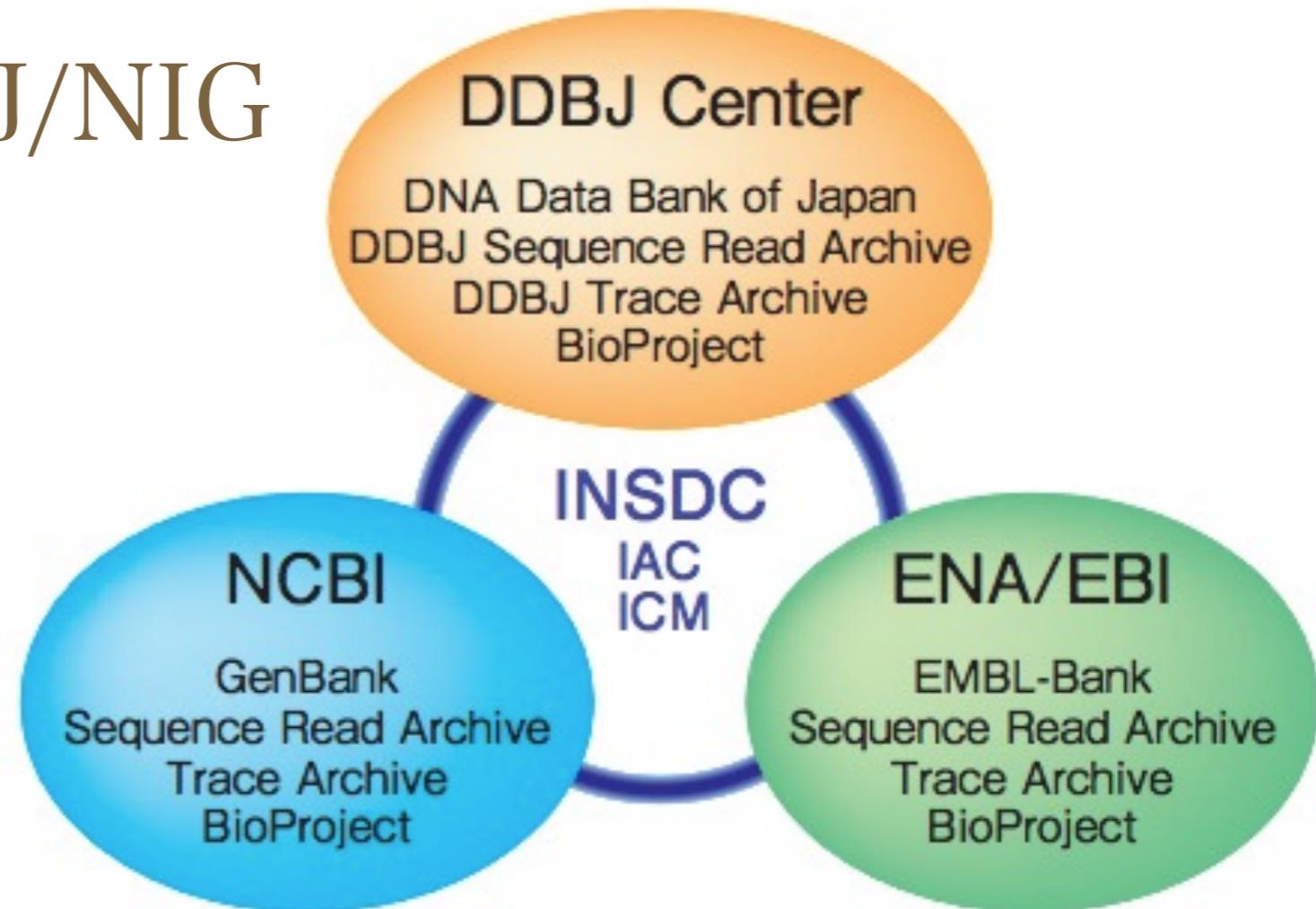


DDBJ は INSDC の一員



International Nucleotide Sequence Databank Collaboration

- USA: GenBank/NCBI
- EU: ENA/EBI
- Japan: DDBJ/NIG



IAC: International Advisory Committee

ICM: International Collaborative Meeting

INSDC personnel, 139 in all



DDBJ (from Release note 92)

Jun Mashima, Hideo Aono, Yuji Ashizawa, Yukino Dobashi, Mayumi Ejima, Masahiro Fujimoto, Asami Fukuda, Tomohiro Hirai, Fumie Hirata, Naofumi Ishikawa, Toshikazu Katsumata, Chiharu Kawagoe, Shingo Kawahara, Yuichi Kodama, Junko Kohira, Takehide Kosuge, Kyungbum Lee, Mika Maki, Kimiko Mimura, Takeshi Moriyama, Yoshihisa Munakata, Naoko Murakata, Keiichi Nagai, Yoshihisa Okido, Yoshihiro Okuda, Katsunaga Sakai, Makoto Sato, Yoshihiro Serizawa, Aimi Shiida, Yukie Shinyama, Rie Sugita, Kimiko Suzuki, Daisuke Takagi, Daisuke Takai, Haru Tsutsui, Koji Watanabe, Tomohiko Yasuda, Shigeru Yatsuzuka, Emi Yokoyama, Eli Kaminuma, Osamu Ogasawara, Kosaku Okubo, Yoshihisa Takagi, and Yasukazu Nakamura

ENA (from Release note 115)

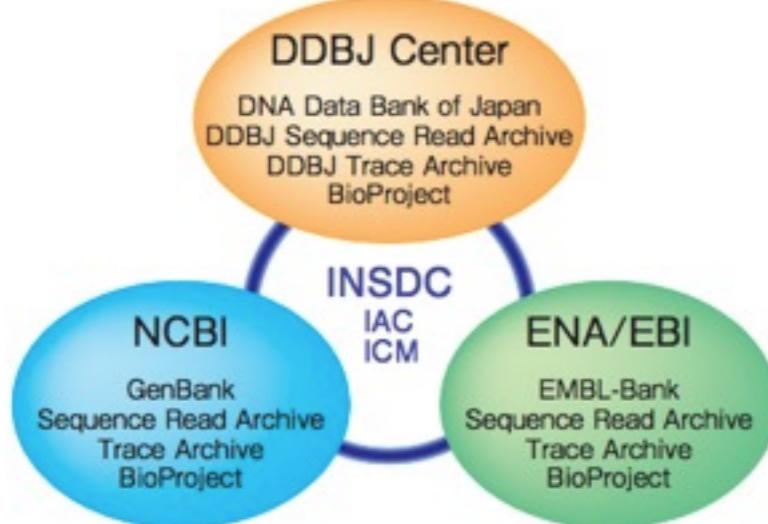
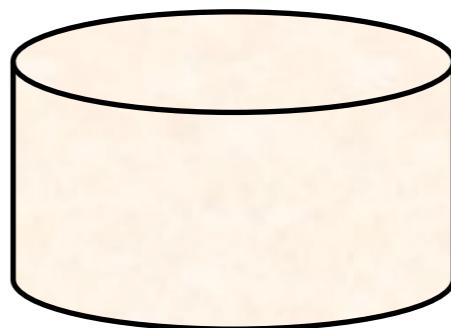
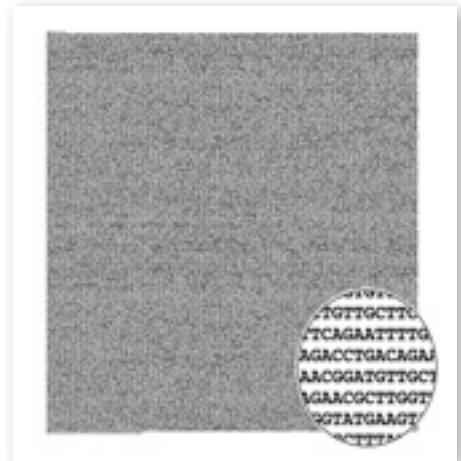
Blaise Alako, Clara Amid, Lawrence Bower, Ana Cerdeno-Taraga, Iain Cleland, Richard Gibson, Neil Goodgame, Petra ten Hoopen, Mikyung Jang, Simon Kay, Rasko Leinonen, Xin Liu, Arnaud Oisel, Rodrigo Lopez, Hamish McWilliam, Nima Pakseresht, Sheila Plaister, Rajesh Radhakrishnan, Kathy Reddy, Stephane Riviere, Marc Rossello, Nicole Silvester, Dmitriy Smirnov, Ana Luisa Toribio, Daniel Vaughan, Vadim Zalunin and Guy Cochrane

GenBank (from Release note 195)

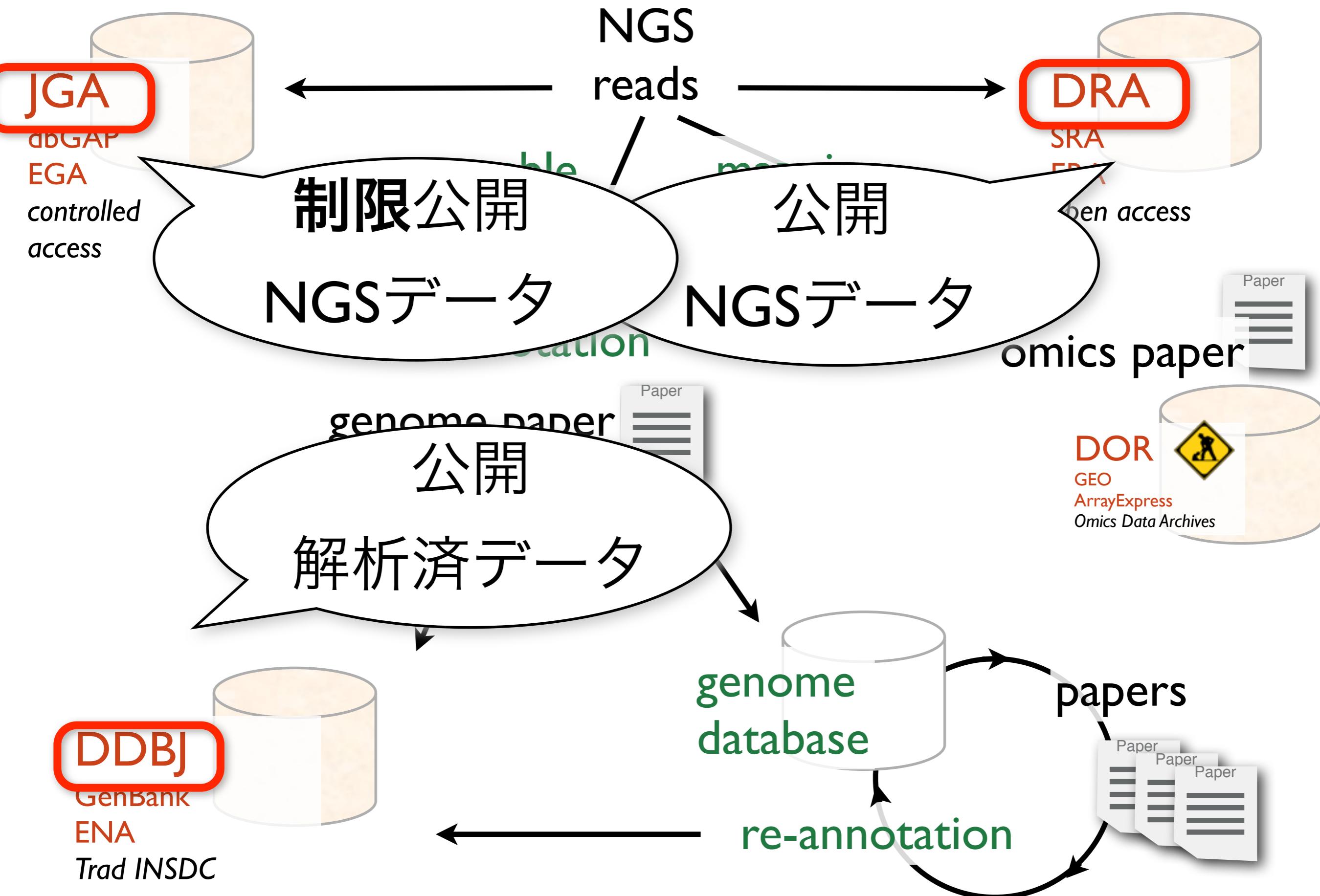
Mark Cavanaugh, Ilene Mizrachi, Yiming Bao, Michael Baxter, Lori Black, Larissa Brown, Vincent Calhoun, Larry Chlumsky, Karen Clark, Jianli Dai, Michel Eschenbrenner, Irene Fang, Michael Fetchko, Linda Frisse, Andrea Gocke, Anjanette Johnston, Mark Landree, Jason Lowry, Suzanne Mate, Richard McVeigh, DeAnne Olsen Cravaritis, Leigh Riley, Susan Schafer, Beverly Underwood, Melissa Wright, Linda Yankie, Serge Bazhin, Evgueni Belyi, Colleen Bollin, Mark Cavanaugh, Yoon Choi, Ilya Dondoshansky, J. Bradley Holmes, WonHee Jang, Jonathan Kans, Leonid Khotomliansky, Michael Kimelman, Michael Kornbluh, Jim Ostell, Denis Sinyakov, Karl Sirotkin, Vladimir Soussov, Elena Starchenko, Hanzhen Sun, Tatiana Tatusova, Lukas Wagner, Eugene Yaschenko, Sergey Zhdanov, Slava Khotomliansky, Igor Lozitskiy, Craig Oakley, Eugene Semenov, Ben Slade, Constantin Vasilyev, Peter Cooper, Hanguan Liu, Wayne Matten, Scott McGinnis, Rana Morris, Steve Pechous, Monica Romiti, Eric Sayers, Tao Tao, Majda Valjavec-Gratian and David Lipman

塩基配列データバンクとは

- 全世界で解読された塩基配列情報を
査定して受入れ
- データベースに蓄積し
- 公開して共有する



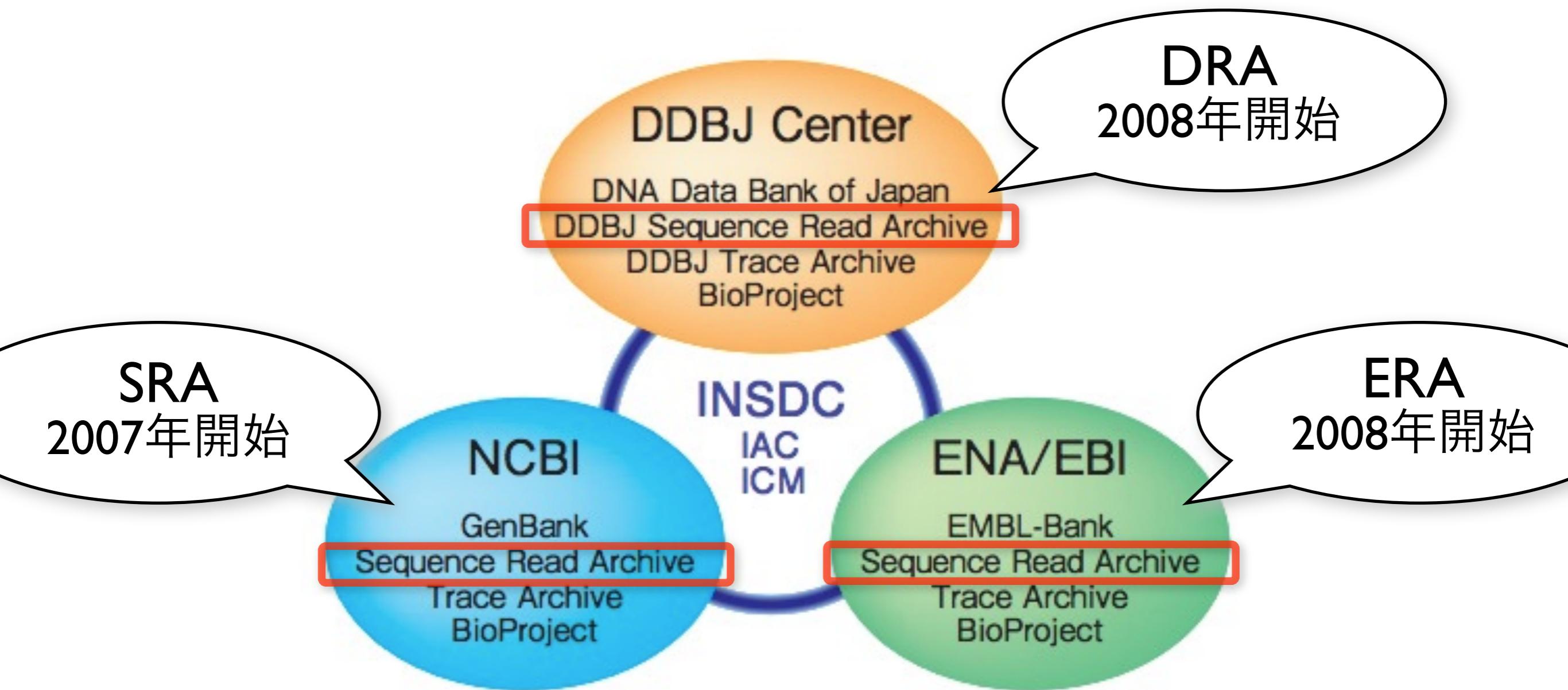
The ecosystem of sequence data



DDBJ Sequence Read Archive (DRA)



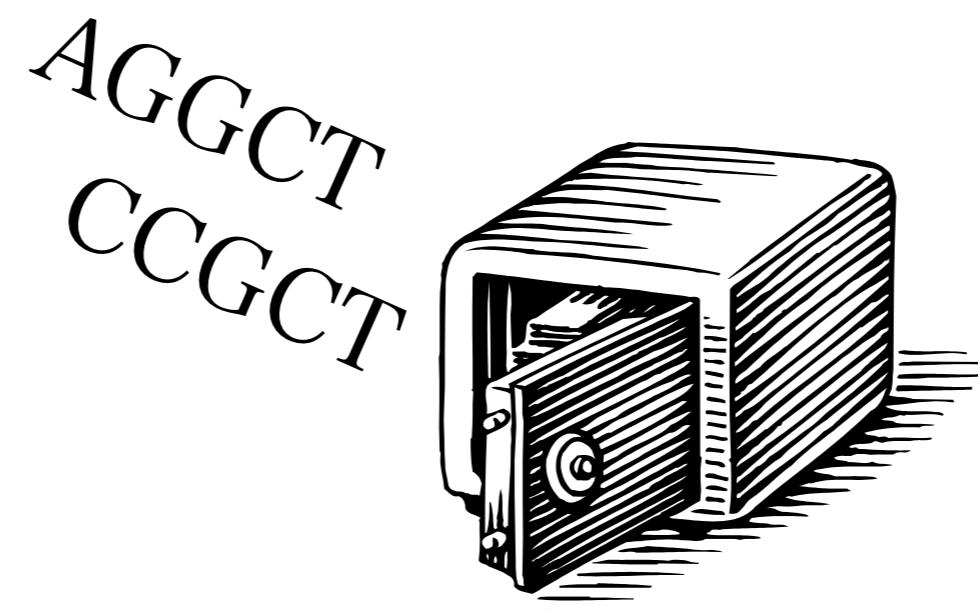
新世代シークエンサから出力される配列や
アライメントデータを登録・公開



INSDC: International Nucleotide Sequence Databank Collaboration

新型シーケンサーデータ保管庫 (DRA)

新型シーケンサーからのデータをDDBJが保管することで、
ユーザがデータを保管するために必要となる
莫大なハードディスクの容量を肩代わりします！
そのかわり、データを公開して良い時期になつたら
公開し、共有・再利用できるように、提供します。



DRA ウェブサイト ⇒ [DRA] で検索



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

登録関係情報

Sequence Read Archive

Login & Submit | Databases | English | Contact

Google™ カスタム検索

Home Submission ▾ Search Download ▾ Pipeline About

解析パイプライン

DDBJ Sequence Read Archive (DRA) は, Roche 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD® System などの次世代シークエンサからの出力データを登録するデータベースです。また、DDBJ は International Nucleotide Sequence Database Collaboration (INSDC) のメンバーであり、International 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センサーによるデータを登録する場合は、DDBJ Sequence Read Archive にご登録ください。

データ検索 データ取得

検索

データをキーワード、生物名、シー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見る

DRA: 新型シーケンサデータを保存・共有



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) との国際連携のもと、運営されています。従来のキャビラリ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。



Released Data					
Accession	Study Title	Organism(s)	Center Name	Release Date	
DRA000001	Whole genome sequencing of <i>Bacillus subtilis</i> subsp. <i>natto</i> BEST195	<i>Bacillus subtilis</i> subsp. <i>natto</i>	KEIO	2010-03-26	Reset
DRA000002	whole genome resequencing of <i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 168	<i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 168	KEIO	2010-03-26	Reset
DRA000010	Whole genome shotgun sequences of <i>Oryza sativa</i> japonica variety, Koshihikari	<i>Oryza sativa</i> Japonica Group	NIAS	2010-03-31	Reset
DRA000030	Whole-genome DNA methylation analysis in human breast cancer cell lines using MeDIP-seq	<i>Homo sapiens</i>	KUGSIPS	2010-03-03	Reset
DRA000039	genetic variation detected in 206 <i>klebsiella pneumoniae</i> plasmids	<i>Klebsiella pneumoniae</i>	WMC	2009-12-14	Reset
DRA000067	<i>B. anthracis</i> BA103 genome analysis	<i>Bacillus anthracis</i>	NED	2010-04-22	Reset
DRA000068	<i>B. anthracis</i> BA104 genome analysis	<i>Bacillus anthracis</i>	NED	2010-04-22	Reset
DRA000099	Whole SNPs analysis of ciprofloxacin resistance among <i>B. anthracis</i> strains	<i>Bacillus anthracis</i>	NED	2010-04-22	Reset
DRA000170	Whole SNPs analysis of ciprofloxacin resistance among <i>B. anthracis</i> strains	<i>Bacillus anthracis</i>	NED	2010-04-22	Reset
DRA000155	CAGE analysis of whole adult brain and whole embryo rat transcriptome	<i>Rattus norvegicus</i>	RIKEN_OSC	2010-03-17	Reset
DRA000169	Linking new promoters to functional transcripts in small samples with nanoCAGE and CAGEscan	<i>Homo Sapiens</i>	RIKEN_OSC	2010-06-08	Reset
DRA000205	A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness	<i>Homo Sapiens</i>	RIKEN_OSC	2010-07-23	Reset
DRA000220	Whole genome sequencing of <i>Oryzias latipes</i> H-9R	<i>Oryzias latipes</i>	KEIO-SM	2010-08-16	Reset
SRA002052	Toxoplasma gondii transcript sequencing project	<i>Toxoplasma gondii</i>	UT-MGS	2009-07-03	Reset
SRA002053	<i>Glossina morsitans</i> transcript sequencing project	<i>Glossina morsitans</i>	UT-MGS	2009-07-03	Reset
SRA002054	<i>Glossina morsitans</i> transcript sequencing project	<i>Glossina morsitans</i>	UT-MGS	2009-06-25	Reset
SRA002055	<i>Anopheles stephensi</i> transcript sequencing project	<i>Anopheles stephensi</i>	UT-MGS	2009-07-01	Reset
SRA002056	<i>Cryptosporidium parvum</i> transcript sequencing project	<i>Cryptosporidium parvum</i>	UT-MGS	2009-07-01	Reset
SRA002057	<i>Plasmodium yoelii</i> transcript sequencing project	<i>Plasmodium yoelii</i>	UT-MGS	2009-09-22	Reset

研究者

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 サイト各箇所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 command:

ftp://www.ddbj.nig.ac.jp/ftphome/DRX/104424-a.html

A new directory, "patient", was made under "ddbj_database".
Now, all of patient value and genome data for JPB and KIPB are included in the new "patient" directory.

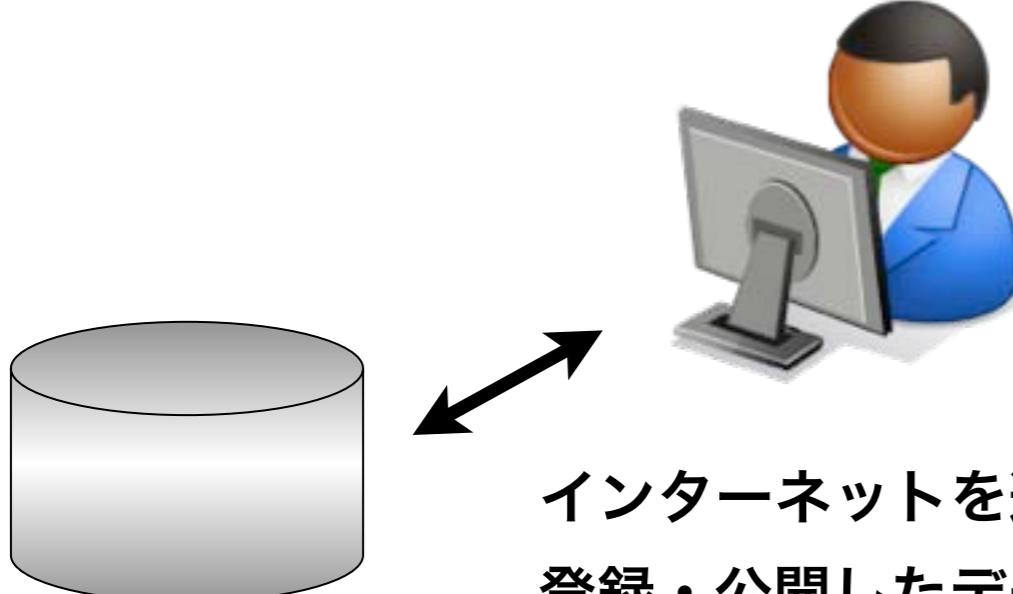
For details, please read the README.TXT in this directory.

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(DNA)
- ftp://ftp.ddbj.nig.ac.jp/ddbj_database/seq/ddbj/
- ftp://ftp.ddbj.nig.ac.jp/ddbj_database/seq/ddbjnew/
- DDBJ-XML
- ftp://ftp.ddbj.nig.ac.jp/ddbj_database/ddbj/xml/medent_current/
- ftp://ftp.ddbj.nig.ac.jp/ddbj_database/ddbjnew/xml/medent_current/
- The last modified on Fri Feb 26 2010.

1階層上のディレクトリへ

07/03/2010 01:17午後 ディレクトリ DRAB19
07/03/2010 01:17午後 ディレクトリ DRAB18
07/03/2010 01:19午後 ディレクトリ DRAB17
07/03/2010 01:19午後 ディレクトリ DRAB16
07/03/2010 01:19午後 ディレクトリ DRAB15
07/03/2010 01:19午後 ディレクトリ DRAB14
07/03/2010 01:19午後 ディレクトリ DRAB13
07/03/2010 01:19午後 ディレクトリ DRAB12
07/03/2010 01:19午後 ディレクトリ DRAB11



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



454 (Roche)



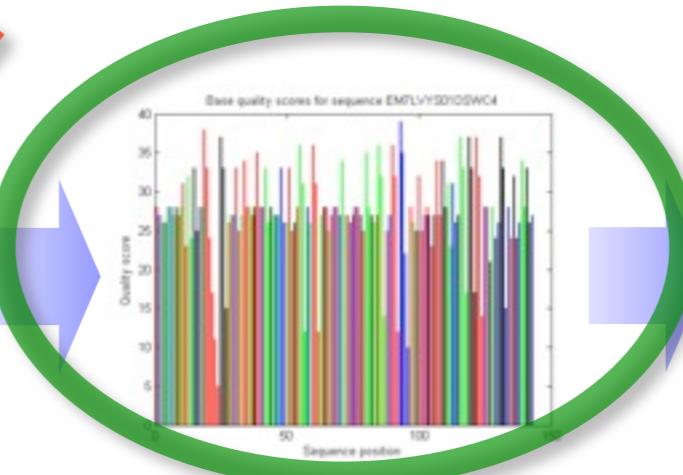
Solexa (Illumina)



SOLiD (ABI) etc

~~配列のみのデータ~~

画像データ



1次データ ベースコール Quality etc

ランデータ

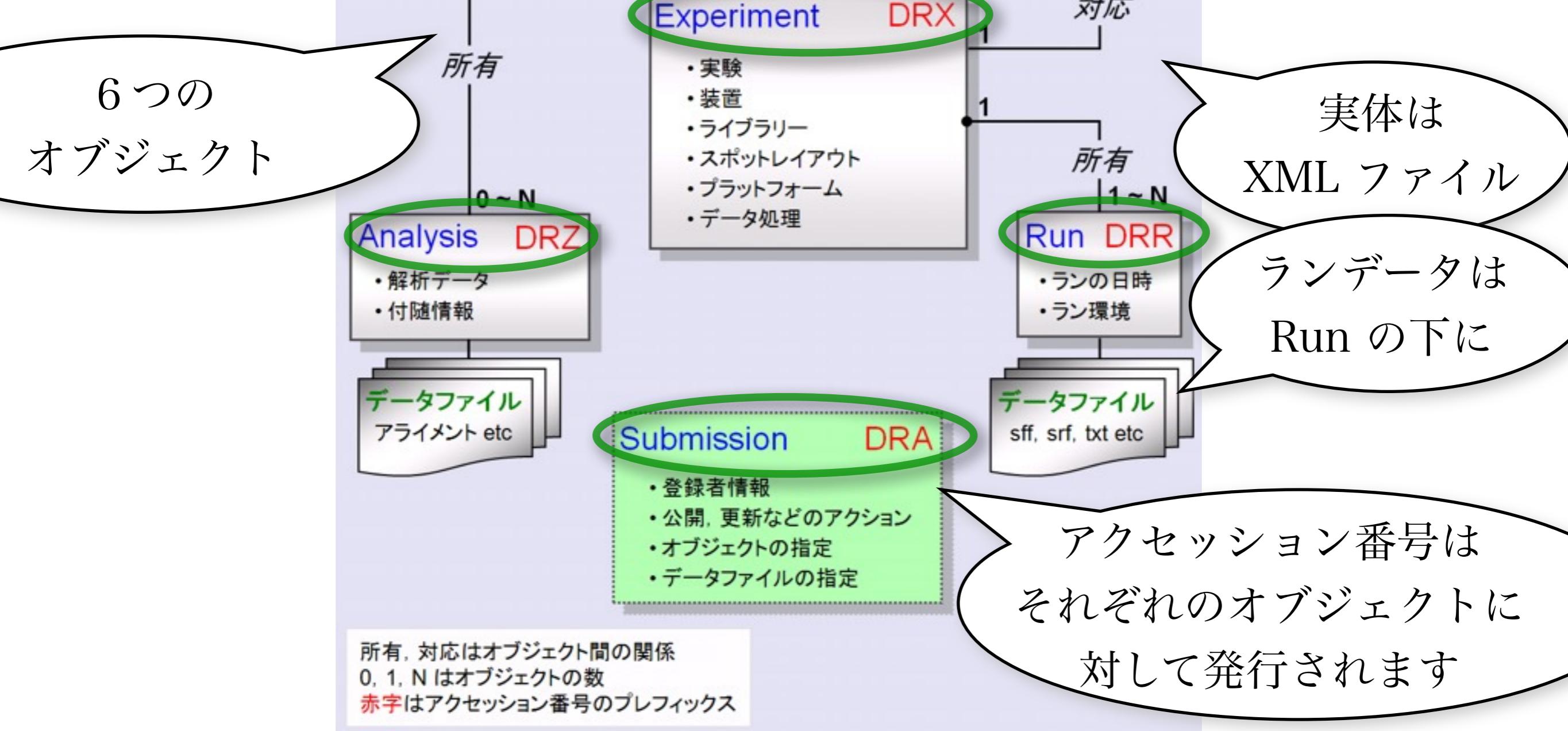
@SRR001654.1 9460:7:1:830:763 length=36
GTCAATATTAATCATACCAATATACTCAAAAAATAA
+SRR001654.1 9460:7:1:830:763 length=36
I+-&*4)%+5'#%/)&\$%\$#%"#&%'%"\$%#%%"!"
@SRR001654.2 9460:7:1:402:781 length=36
GGTCTAAAAAGCAAAATTCAAGTCTTCAAAATAATT
+SRR001654.2 9460:7:1:402:781 length=36
II+(%\$+%'&+*-0+/*("%&+"*&"(*\$""#%&%\$
@SRR001654.3 9460:7:1:433:775 length=36
GTGCTTTTTTTTCCAGGAAGTTGTCTCCTCTATC
+SRR001654.3 9460:7:1:433:775 length=36
II3DI>IIIH%&%&%+"%,\$&\$&%#
fastq データ

塩基配列 と Quality Value

メタデータ

(データに関する情報：実験手法、解析方法 etc)

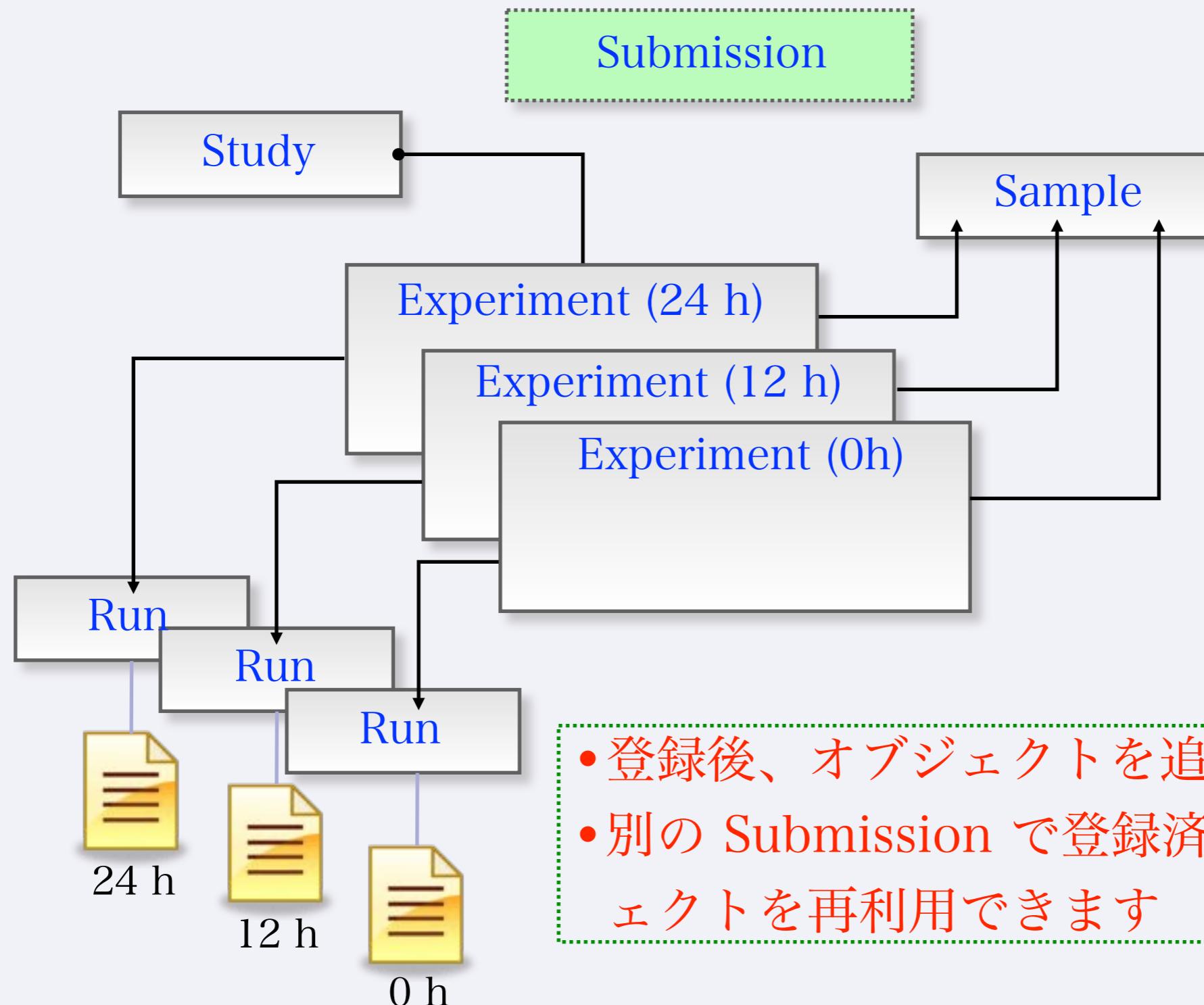
メタデータ



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



【実習】 DRAsearch で検索してみよう



The screenshot shows the DDBJ Sequence Read Archive (DRA) homepage. At the top, there is a navigation bar with links for Home, Submission, Search (which is highlighted with a red circle), Download, Pipeline, and About DRA. To the right of the navigation bar is a search bar containing the text "Google™ カスタム検索" and a magnifying glass icon. Below the navigation bar, there is a brief description of what DRA is and how it relates to other international sequence databases. On the left side of the main content area, there is a large circular icon with a magnifying glass symbol, labeled "検索" (Search). Below this icon, there is Japanese text: "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". To the right of this are two smaller rectangular boxes: one for "登録" (Registration) showing a stack of discs with an upward arrow, and another for "動画マニュアル" (Video Manual) showing a video camera icon.

[DRA] で google 検索 ⇒

DDBJ Sequence Read Archive のサイトを開く

DRAsearch: *A. thaliana* を選択

DRA Search trace.ddbj.nig.ac.jp/DRASearch/ その他のブックマーク

W LSDB E P R S Tech Papers Prp

DRASearch Send Feedback Search Home DRA Home

Accession :

Organism : Ara StudyType :

CenterName : Arabidopsis arenosa

Keyword : Arabidopsis arenosa x Arabidopsis thaliana

Show 20 results Clear

Data Last Update 2013-08-28
WebSite Last Update 2013-08-28

Statistics [Arabidopsis thaliana](#)

Released Experiment Type Submission Study Experiment Sample Run

Arabis alpina	Arachis duranensis	Arachis hypogaea	437191	481853
---------------	--------------------	------------------	--------	--------

Organism

#	Organism Name	Study
1	Homo sapiens	2015
2	Mus musculus	1319
3	unidentified	875
4	Drosophila melanogaster	510
5	Caenorhabditis elegans	282

Study Type

#	Study Type	Study
1	Whole Genome Sequencing	13228
2	Other	3853
3	Metagenomics	1781
4	Transcriptome Sequencing	1759
5	Transcriptome Analysis	1217

Center Name [All List]

#	Center Name	Study
1	BioProject	6425
2	GEO	3550
3	JGI	3133
4	UMIGS	1637
5	ICVIT	1161

#	Organism Name	Study	#	Study Type	Study	#	Center Name	Study
1	Homo sapiens	2015	1	Whole Genome Sequencing	13228	1	BioProject	6425
2	Mus musculus	1319	2	Other	3853	2	GEO	3550
3	unidentified	875	3	Metagenomics	1781	3	JGI	3133
4	Drosophila melanogaster	510	4	Transcriptome Sequencing	1759	4	UMIGS	1637
5	Caenorhabditis elegans	282	5	Transcriptome Analysis	1217	5	ICVIT	1161

DRAsearch: StudyType を選択

DRA Search trace.ddbj.nig.ac.jp/DRASearch/ その他のブックマーク

W LSDB E P R S Tech Papers Prp

Send Feedback Search Home DRA Home

Accession :

Organism : Arabidopsis thaliana StudyType :

CenterName : Platform :

Keyword :

Show 20 records Sort by Study Search Clear

Statistics

Released Entries

Type	Count
Submission	151844
Study	23412
Experiment	354870
Sample	437191
Run	481853

Organism

#	Organism Name	Study
1	Homo sapiens	2015
2	Mus musculus	1319
3	unidentified	875
4	Drosophila melanogaster	510
5	Caeprababditis elegans	282

Study

#	S	Study
1	Whole Genome Sequencing	13228
2	Other	3853
3	Metagenomics	1781
4	Transcriptome Sequencing	1759
5	Transcriptome Analysis	1217

Center Name [All List]

Center Name	Study
BioProject	6425
GEO	3550
JGI	3133
UMIGS	1637
ICVIT	1161

✓

- 16S pyrosequencing
- Amplicon sequencing
- Cancer Genomics
- Epigenetics
- Exome Sequencing
- Gene Regulation Study
- genome partial sequencing
- genome-wide SNP discovery
- Metagenomics
- Metatranscriptomics
- multi-isolate
- Other
- pooled clone
- Pooled Clone Sequencing
- Population Genomics
- Resequencing
- RNASeq
- small RNA
- Synthetic Genomics
- Transcriptome Analysis
- Transcriptome Sequencing**
- vector insertion site analysis
- Whole genome bisulfite sequencing
- Whole Genome Sequencing

Data Last Update 2013-08-28
WebSite Last Update 2013-08-28

DRAsearch: 検索結果リスト



Result List - DRA Search

trace.ddbj.nig.ac.jp/DRASearch/query?organism=Arabidopsis+thaliana&study_type=Transcriptome+Sequencing¢er_n... ☆

W LSDB E P R S Tech Papers Prp その他のブックマーク

DRASearch Send Feedback Search Home DRA Home

Accession :

Organism : StudyType :

CenterName : Platform :

Keyword :

Show 20 records Sort by Study Search Clear

Search Results (76 studies) << < 1 / 4 Page > >>

#	STUDY	SUBMISSION	STUDY_TITLE	STUDY_TYPE	ORGANISM	BASES	SUBMITTED	CENTER_NAME
1	SRP000551	SRA000285	Highly integrated epigenome maps in Arabidopsis - small RNA sequencing	Transcriptome Sequencing	Arabidopsis thaliana	1.5G	2009-02-07	GEO
2	SRP000568	SRA000286	Highly integrated epigenome maps in Arabidopsis - transcriptome sequencing	Transcriptome Sequencing	Arabidopsis thaliana	12.7G	2009-02-14	GEO
3	SRP000615	SRA008180	Comparison of next generation sequencing technologies for transcriptome characterization	Transcriptome Sequencing	Eschscholzia californica Persea americana	118.2M	2009-03-05	Penn State University
4	SRP000622	SRA008236	A link between RNA metabolism and silencing affecting Arabidopsis development	Transcriptome Sequencing	Arabidopsis thaliana	4.1G	2009-03-11	GEO
5	SRP000625	SRA008240	AGO1-dependent Small RNA in Arabidopsis Identified by High-Throughput Sequencing	Transcriptome Sequencing	Arabidopsis thaliana	829.7M	2009-03-01	GEO
6	SRP000713	SRA008384	Novel microRNA-target RNA pairs revealed by Parallel analysis of RNA ends	Transcriptome Sequencing	Arabidopsis thaliana	425.1M	2009-04-19	GEO
7	SRP000730	SRA008413	Small RNAs in Arabidopsis hybrid siliques	Transcriptome Sequencing	Arabidopsis thaliana	167.2M	2009-04-15	GEO
8	SRP000935	SRA009031	Transcriptome-wide map of alternative splicing in Arabidopsis	Transcriptome Sequencing	Arabidopsis thaliana	12.5G	2009-07-07	OSU-CGRB

Computational and Analytical Framework for

DRAsearch: Study Detail

Result List – DRA Search × SRP000935 – DRA Search × SRX006191 – DRA Search ×

trace.ddbj.nig.ac.jp/DRASearch/study?acc=SRP000935

W LSDB E P R S Tech Papers Prp その他のブックマーク

DRASearch Send Feedback Search Home DRA Home

SRP000935

Study Detail

Title	Transcriptome-wide map of alternative splicing in Arabidopsis
Study Type	Transcriptome Sequencing
Abstract	placeholder
Description	Alternative splicing was profiled in <i>Arabidopsis thaliana</i> 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 (Oregon St Univ - Center for Genome Research & Biocomputing)

Navigation

Submission	SRA009031	
Experiment	SRX006191	
	SRX006192	
	SRX006193	
	SRX006681	
	SRX006682	
	SRX006688	
	SRX006690	
	SRX006692	
	SRX006704	
	SRX179490	
	SRX179491	
	SRX179492	
	SRX179493	
	SRX179494	
	SRX179495	
Sample	SRS004095	
	SRS004097	
	SRS004098	
	SRS004099	
	SRS004100	
	SRS004101	
	SRS004102	
	SRS004103	
	SRS004104	

DRAsearch: Experiment Detail

Result List – DRA Search × SRP000935 – DRA Search × SRX006191 – DRA Search ×

trace.ddbj.nig.ac.jp/DRASearch/experiment?acc=SRX006191

W LSDB E P R S Tech Papers Prp その他のブックマーク

DRASearch Send Feedback Search Home DRA Home

SRX006191 FASTQ SRALite

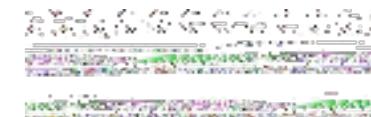
Experiment Detail

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

Navigation

Submission	SRA009031	FTP
Study	SRP000935	
Sample	SRS004095	
Run	SRR018346	FASTQ SRALite

例: この SRA データを使った二次論文



SRP000935 pubmed



ウェブ 画像検索 地図 ショッピング もっと見る

4 件 (0.19 秒)

[Determinants of Exon-Level Evolutionary Rates in *Arabidopsis* Species](#)

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 Schmidt's lab (SRP007763). These RNA-seq data cover seven conditions, including abiotic stresses (cold, drought, heat, highlig

Evol Bioinform Online. 2012; 8: 389–415.

PMCID: PMC3399485

Published online 2012 July 4. doi: [10.4137/EBO.S9743](https://doi.org/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}

¹Graduate Institute of Life Sciences, National Defense Medical Center, 114 Taiwan

²Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, 305 Taiwan

³Department of Life Sciences, National Chiao-Tung University, Hsinchu, 300 Taiwan

⁴Department 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](https://doi.org/10.3389/fgene.2012.00239)

[Estimation of Gene Expression at Isoform Level from mRNA-Seq Data by Bayesian Hierarchical Modeling](#)

www.ncbi.nlm.nih.gov > ... > Front Genet > v.3; 2012 ▾ このページを訳す

M Bhattacharjee 著 - 2012 - 関連記事

2012/11/27 - The dataset was downloaded from NCBI GEO website (SRP000935). Pooled data contains [PubMed]; Bhattacharjee M, Nelson P. S., Arjas E. (2004). Bayesian integrated functional analysis of gene expression data across multiple tissues. *Statistica Sinica* 14: 1111–1127.

Nelson P. S., Arjas E. (2004). Bayesian integrated functional analysis of gene expression data across multiple tissues. *Statistica Sinica* 14: 1111–1127.

[7075 - DNAexus - SRA](#)

sra.dnanexus.com/?page=7075&result_type... ▾ このページを訳す

SRR019209, Oregon St Univ - Center for Genome Research & Biotechnology, USA

SRX006688 - Arabidopsis thaliana - **SRP000935**, SRA009031. 1

Oregon St Univ ... pubmed(21393572)» · SRR099327, Stanford

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. Davuluri³

¹Department of Statistics, University of Pune, Pune, India

²Department of Mathematics and Statistics, University of Hyderabad, Hyderabad, India

³Center 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:

FAQ, Manual, 動画マニュアルをどうぞ

The screenshot shows a web browser window with the following details:

- Address Bar:** trace.ddbj.nig.ac.jp/dra/index.html
- Toolbar:** Detail view, DDBJ Sequence Read Arch, back, forward, search, etc.
- Menu Bar:** W, LSDB, E, P, R, S, Tech, Papers, Prp, その他ブックマーク
- Header:** DDBJ (DNA Data Bank of Japan), Sequence Read Archive, Login & Submit, Databases, English, Contact, Google カスタム検索, 検索ボタン
- Navigation:** Home, Submission (highlighted), Search, Download, Pipeline, About DRA
- Left Sidebar (Submission):**
 - Submission
 - DDBJ Sequence Read Archive (説明)
 - Metadata
 - Data File
 - Example
 - FAQ
 - Manual
 - Video Tutorial
 - Login & Submit
- Content Area:**
 - 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見る
- Bottom Links:** Databases (Nucleotide Sequence Database, Sequence Read Archive, Trace Archive, Omics Archive, BioProject, BioSample), Resources (getentry, ARSA, TXSearch, BLAST, Vector Screening System, ClustalW, Read Annotation Pipeline), DDBJ Information (DDBJ RSS, DDBJ on Twitter, DDBJ on Youtube, DDBJ Web Magazine, DDBJing, DDBJ FTP Site).

A red circle highlights the "動画マニュアル" (Video Manual) link under the "Submission" section, which is described as a video explaining how to use and submit data to DRA.

ネットワークを介して計算機を使い倒す



遺伝研

スーパーコンピュータ



スパコン利用申請はこちら



- [遺伝研 スーパーコンピュータ] で検索

検索...

NIG SuperComputer

SuperComputer Facilities of National Institute of Genetics

現在地: Home

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)

システム情報

+ システム構成
+ システム使用方法
稼働スケジュール

各種申請・問い合わせ

+ 各種申請
+ 問い合わせ

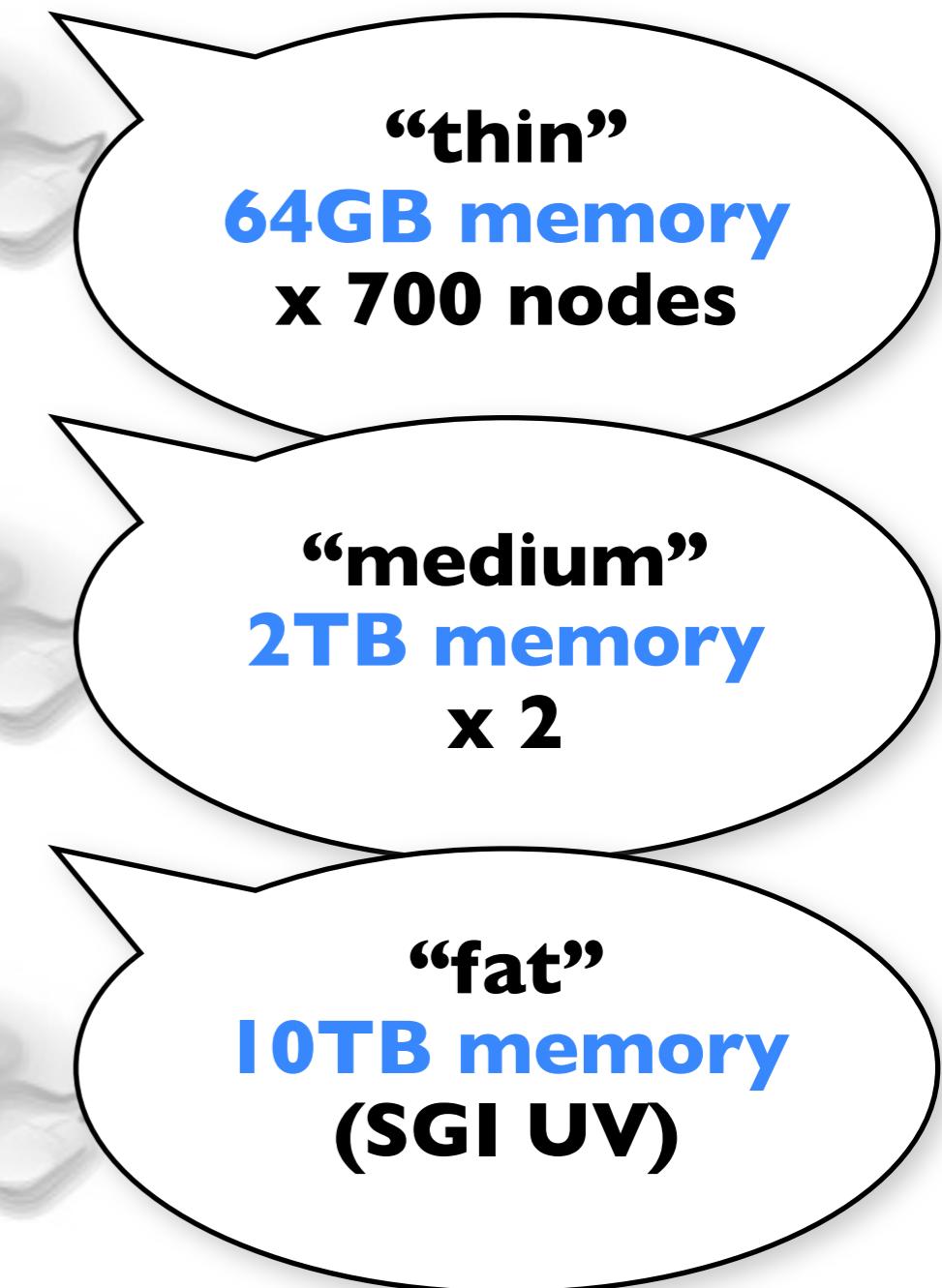
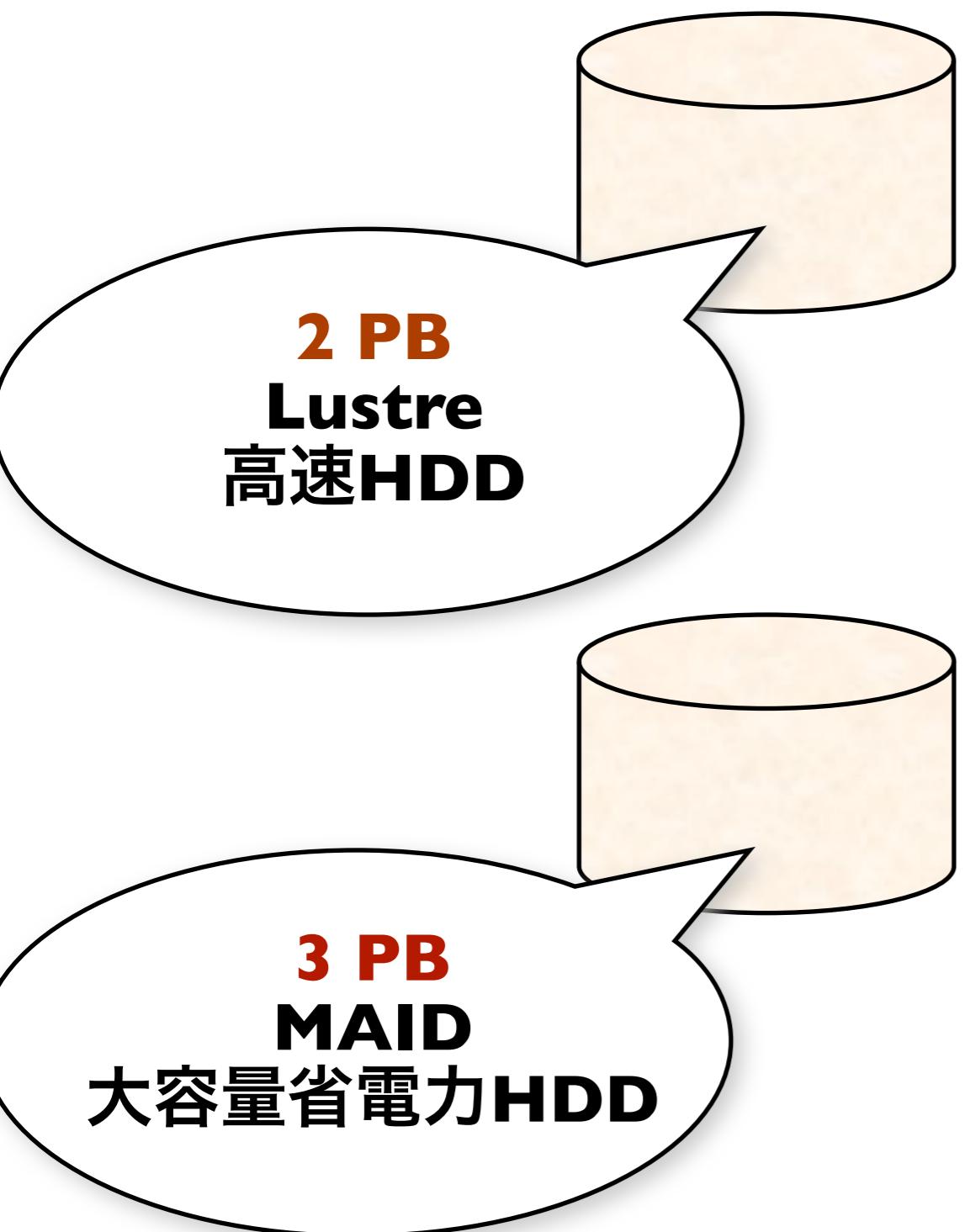
重要なお知らせ

公開日	表題
2013年8月20日	計算機リソース拡張利用申請中止のお知らせ
2013年7月31日	バックアップサービス中止のお知らせ
2013年3月27日	NCBI ftpサイトへのアクセスについて
2013年2月4日	【再依頼】ディスク使用量削減のお願い

新着のお知らせ

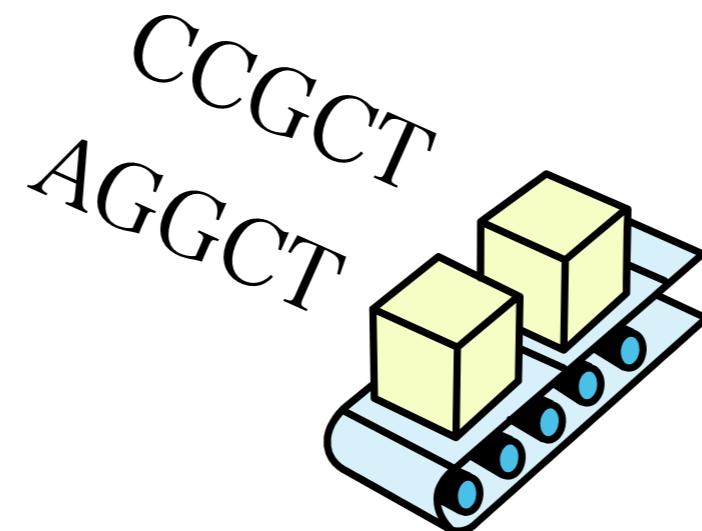
公開日	表題
2013年8月26日	fatノードサービス再開のお知らせ
2013年8月23日	fatノードサービス障害のお知らせ
2013年8月22日	国立遺伝学研究所スーパーコンピュータシステム サービス復旧のお知らせ
2013年8月21日	fatノードサービス再開のお知らせ
2013年8月20日	fatノードサービス障害のお知らせ

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



DDBJ pipeline

情報処理用語でパイプラインとは、さまざまなプログラムを直列に連結し、工場などの流れ作業のようにプログラムを連続して走らせてデータの処理を高速化・定型化する技術です
新型シークエンサーからのデータを、DDBJのスーパーコンピュータを使って解析できるパイプラインを用意し、莫大なデータを解析するための計算機資源を個別に用意しなくてもよく、また、複雑な情報処理技術なしで、各種の解析処理がボタンひとつで簡単に実行できます！



DDBJ Pipeline: クラウド型解析ツール

The screenshot shows the DDBJ website interface. On the left, there's a sidebar with links for Web Magazine, RSS, and Twitter. Below that are logos for DDBJ, INSDC, NCBI, and ENA/EBI. The main content area has a header "DDBJ Service". A red circle highlights the "Search / Analysis" button, which is connected by a red arrow to a separate window titled "検索・解析" (Search & Analysis). This window lists various search tools like getentry, ARSA, TXSearch, BLAST, DDBJ Vector Screening System, and DRA Search. Another red circle highlights the "次世代 Sequence 解析" (Next-generation Sequence Analysis) section at the bottom right, which includes the "DDBJ Read Annotation Pipeline".

DDBJ Service

English

Googleカスタム検索 Search

DDBJ の紹介 利用の手引き レポート・統計 Q and A お問い合わせ

Web Magazine RSSを購読する DDBJ Twitter

DDBJ INSDC NCBI ENA/EBI International Nucleotide Sequence Database Collaboration

登録 検索・解析 Data Submission Search / Analysis

Hot Topics

- 2013.08.09 第26回国際実務者会議 報告
- 2013.08.06 EMBL-Bank 由来 TPA データの不備についてお詫び
- 2013.08.02 サンゴに共生するカッチャウソウ (Symbiodinium)

Maintenance

www.ddbj.nig.ac.jp/searches-j.html

検索・解析

検索・解析 | DDBJ www.ddbj.nig.ac.jp/searches-j.html

W LSDB E P R S Tech Papers Prp

検索・解析

データベース検索

getentry アクセション番号などによるエントリの検索

ARSA 高速なキーワード検索

TXSearch 生物分類データベース検索

BLAST 相同性検索

DDBJ Vector Screening System ベクター配列データベースを対象にした相同性検索サービス

DRA Search DRA に登録されたデータを、キーワード、生物名、シークエンサなどで検索

系統解析

ClustalW 基配列・アミノ酸配列の多重整列と系統樹作成

ゲノム解析

GIB ※一時中断 真正細菌・古細菌・真核生物ゲノムの統合検索

GIB-V ※一時中断 ウィルスゲノムの統合検索

MiGAP 微生物ゲノム配列のアノテーションツール
(ご利用には申請が必要です)

MiGAP-OLD (2012年2月以前の解析結果利用) 微生物ゲノム配列のアノテーションツール
(ご利用時のDBCLSのOpenIDが必要です)

GTPS 共通プロトコルに基づくバクテリアゲノムの再アノテーション

GTOP ゲノム配列からタンパク質の構造へ

次世代 Sequence 解析

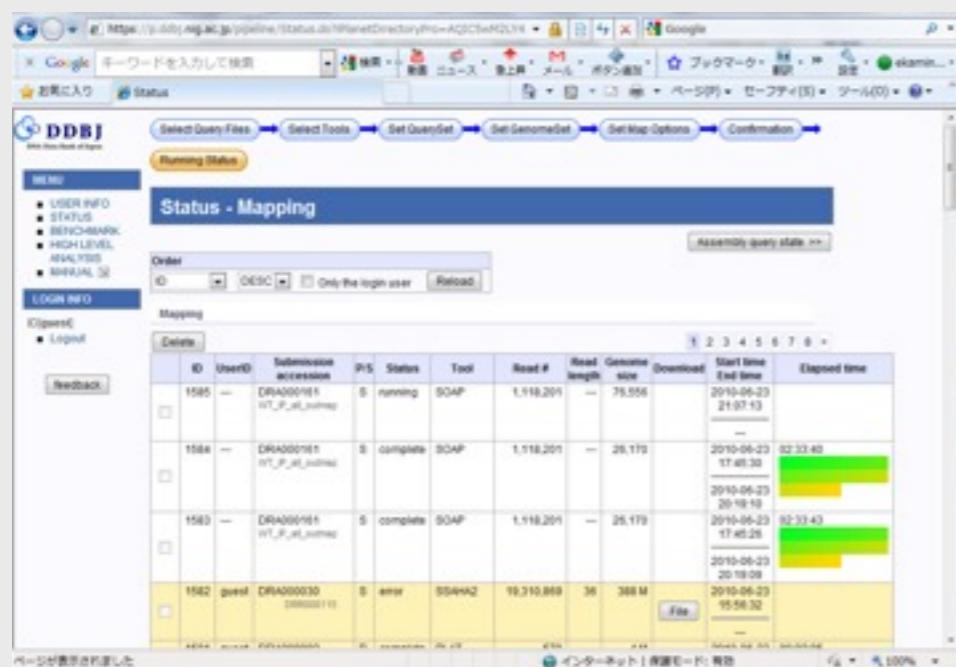
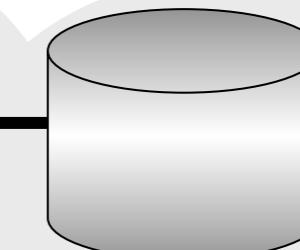
DDBJ Read Annotation Pipeline 次世代シークエンサ出力データの解析システム
(ご利用には申請が必要です)

PAG

DDBJ Pipeline: クラウド型解析ツール



研究者



DDBJ スーパーコンピュータ



計算資源が足りない → DDBJスパコンを使おう
解析技術がない → web ブラウザから自分で解析

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

DDBJ
DNA Data Bank of Japan

ACCOUNT
login ID [yaskaz]
[Logout](#)
[Change password](#)

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

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

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

Select Query Files → Select Tools → Set QuerySet → Set Genome → Set Genomic Features

Running Status

Selecting Tools for Basic Analysis of DDBJ ANNOTATION PIPELINE

BACK **NEXT**

① 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	
BLAT		34	✓				✓						Single-end analysis only	
Mag		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 sensitive, and uses less memory than Bowtie1.	

② 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 severely recommend when performing Velvet, total length of those reads is up to 22G bp. Maximum K-mer value is 64.

良く使われるソフトウェアの特徴

マッピング

BLAT	高速シーケンサー登場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	高速シーケンサー登場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。
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にたいしてインデックスを作成、高速でマッピングする。 Bowtie2は50bp以上に最適化。
TopHat	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ー。
Velvet	高速シーケンサー登場初期に開発された。メモリ消費多め。
Trinity	RNA-Seq配列のアセンブラー。上記3つとともにde bruijn graphというアルゴリズムを使用

DRA pipeline: 比較対象

The screenshot shows the DDBJ DRA pipeline interface. On the left, there's a sidebar with 'ACCOUNT' and 'ANALYSIS' sections, and a 'JOB STATUS' section at the bottom. The 'JOB STATUS' section has a blue bar highlighting 'step1. Preprocessing'. In the main area, there's a flowchart: 'Select Query Files' → 'Select Tools' → 'Set QuerySet'. Below it, a large button says 'Specifying Database of Reference'. Under this button, there's a section titled 'Major genome sets' with a red arrow pointing to the 'Organisms' dropdown which is set to 'Arabidopsis thaliana'. A dropdown menu for 'Genome sets' shows 'TAIR8' checked, along with 'TAIR9' and 'TAIR10'. Another section below shows genome files: 'all.fa' is checked, while 'chr01.fa' through 'chrM.fa' are unchecked.

イネ、マウスなど
解析比較対象となる

配列を多数用意

Major genome sets

Organisms

Oryza sativa japonica

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 mitochondrial

tigr chloroplast

Major genome sets

Organisms

Homo sapiens

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

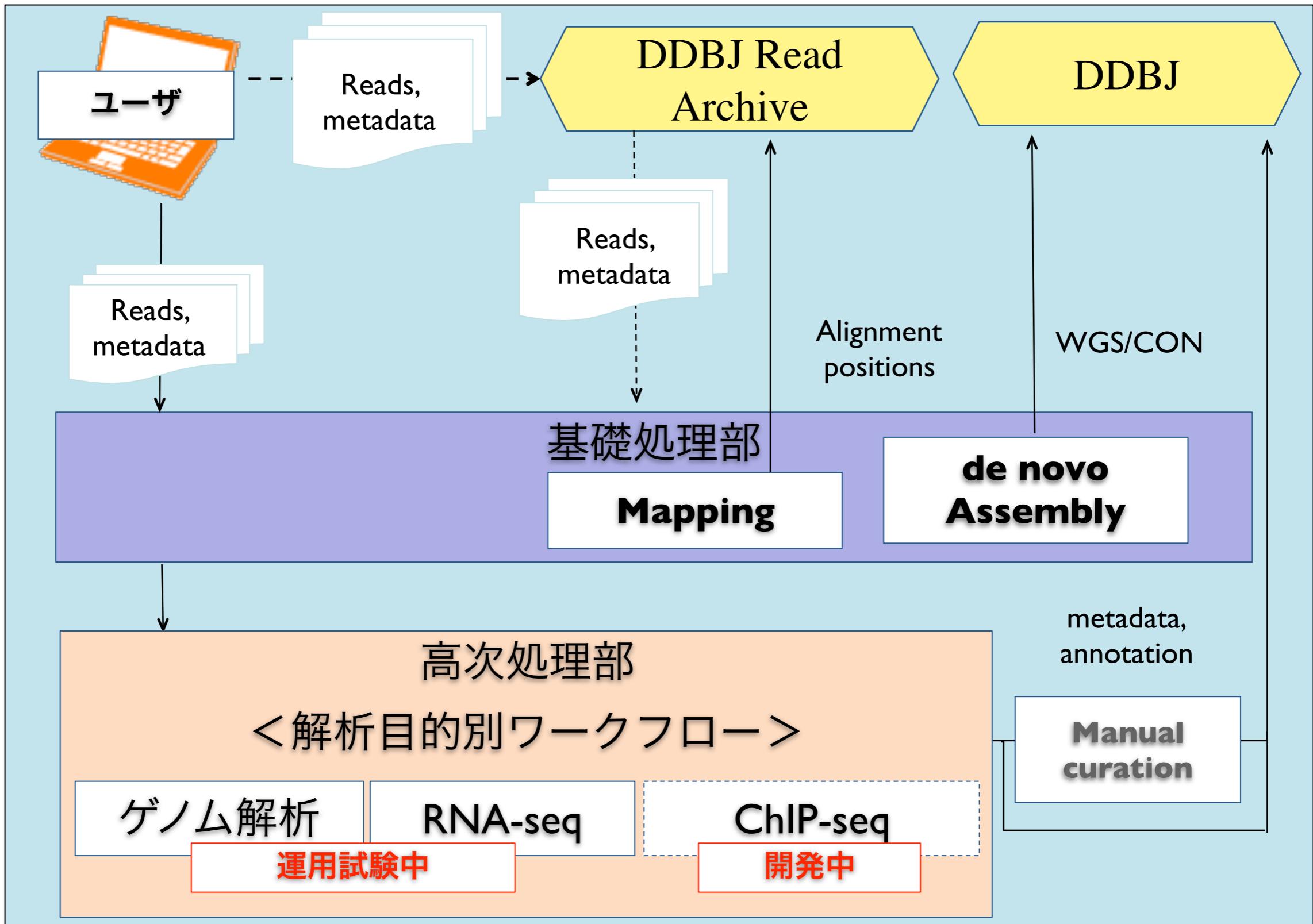
NCBI build 37.1_CRA

NCBI build 37.1_Celera

NCBI build 37.1_GRCh

NCBI build 37.1_HuRef

DDBJ Pipeline 概要



<http://p-galaxy.genes.nig.ac.jp/>

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data', 'Help', and 'User'. On the left, a sidebar titled 'Work Flow' lists various tools under categories like 'SET DATA', 'RNA-SEQ', 'SNP DETECTION', and 'Get Data'. The main content area displays a workflow titled 'WWFSMD? grow noodly appendages...'. This workflow is composed of several interconnected steps: 'Input dataset' (with 'out_file'), 'Filter' (with 'out_file1'), 'Join' (with 'output (interval)' and 'out_file2'), 'Sort' (with 'Sort Query' and 'out_file3'), and 'Select first' (with 'from' and 'out_file3'). Below the workflow, the URL 'usegalaxy.org' is displayed. At the bottom, a note states: 'This project is supported in part by NSF, NHGRI, and the Huck Institutes of the Life Sciences.' A large speech bubble on the right side of the screen contains the Japanese text 'HLA 解析系も こちらで提供'.

ツール オプション ▾

Work Flow

SET DATA

- [import samfile](#)
- [import samfile na](#)
- [import samfile mo](#)

RNA-SEQ

- [Cufflinks preprocessing](#)
- [convert gtf to wig for exp. analysis by Cufflinks](#)

SNP DETECTION

- [uniq BWA sam](#)
- [join SAMfile](#)
- [SAM-to-BAM](#) converts SAM format to BAM format
- [Generate pileup](#) from BAM dataset
- [Filter pileup](#) on coverage and SNPs
- [draw distribution on genome](#)

Get Data

Send Data

ENCODE Tools

Lift-Over

Analyze Data Workflow Shared Data Help User

Hello world! It's running...

To customize this page edit static/welcome.html

WWFSMD?
grow noodly appendages...

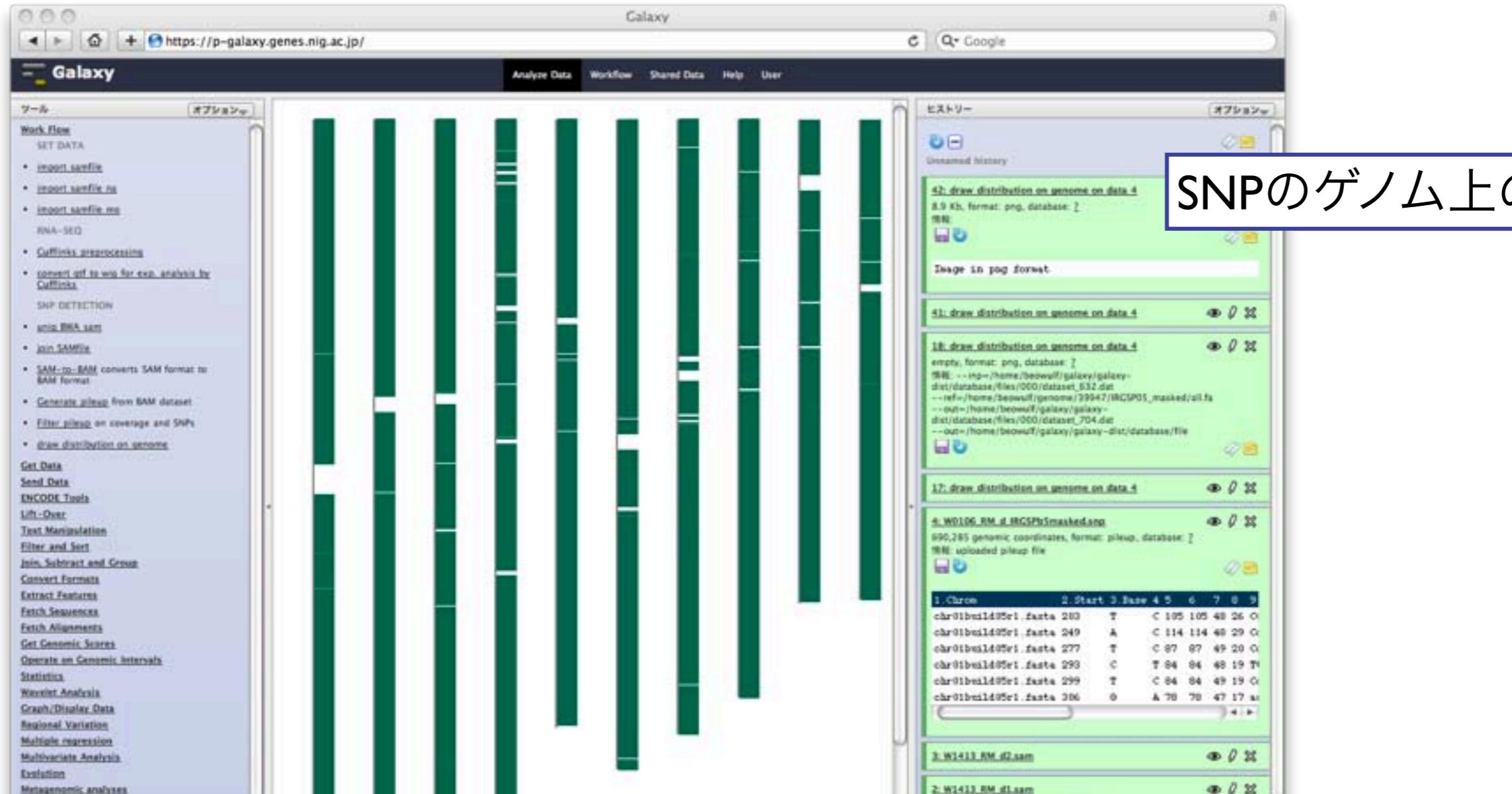
usegalaxy.org

This project is supported in part by [NSF](#), [NHGRI](#), and [the Huck Institutes of the Life Sciences](#).

HLA 解析系も
こちらで提供

インターネット 100%

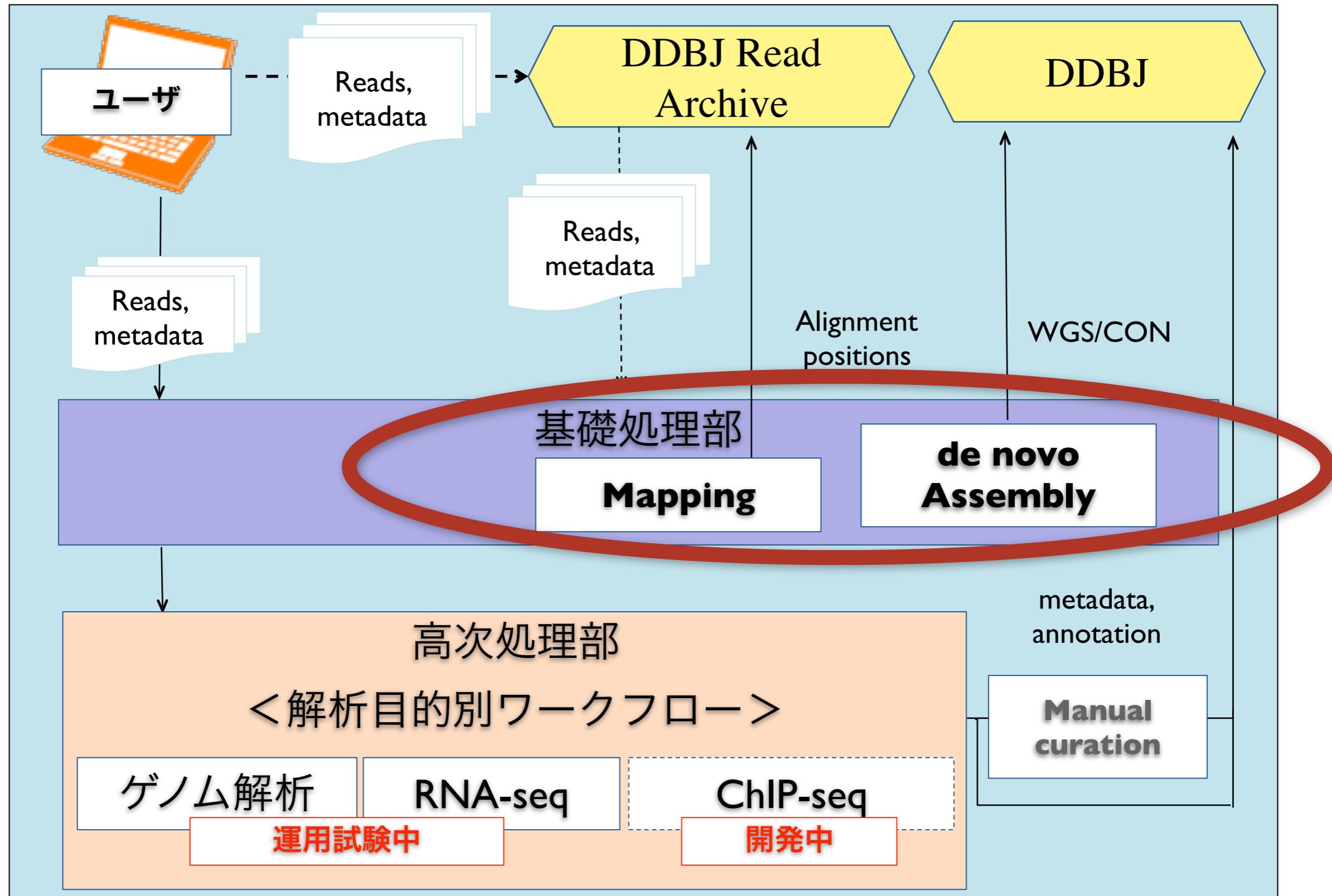
ゲノム解析ワークフローの実行結果例



SNPのゲノム上の分布図

```
chr01build05r1.fasta samtools.snp 203 203 105 + . Note=C/T; Note=consQ:105 SNPQ:105 MaxMapQ:48 readCov:26;
chr01build05r1.fasta samtools.snp 277 277 87 + . Note=C/T; Note=consQ:87 SNPQ:87 MaxMapQ:49 readCov:20;
chr01build05r1.fasta samtools.snp 293 293 84 + . Note=T/C; Note=consQ:84 SNPQ:84 MaxMapQ:48 readCov:19;
chr01build05r1.fasta samtools.snp 299 299 84 + . Note=C/T; Note=consQ:84 SNPQ:84 MaxMapQ:49 readCov:19;
chr01build05r1.fasta samtools.snp 306 306 78 + . Note=A/G; Note=consQ:78 SNPQ:78 MaxMapQ:47 readCov:17;
chr01build05r1.fasta samtools.snp 319 319 78 + . Note=T/C; Note=consQ:78 SNPQ:78 MaxMapQ:49 readCov:17;
chr01build05r1.fasta samtools.snp 598 598 81 + . Note=A/G; Note=consQ:81 SNPQ:81 MaxMapQ:43 readCov:18;
chr01build05r1.fasta samtools.snp 751 751 57 + . Note=A/T; Note=consQ:57 SNPQ:57 MaxMapQ:33 readCov:10;
chr01build05r1.fasta samtools.snp 1027 1027 75 + . Note=A/G; Note=consQ:75 SNPQ:75 MaxMapQ:46 readCov:16;
chr01build05r1.fasta samtools.snp 1068 1068 66 + . Note=T/C; Note=consQ:66 SNPQ:66 MaxMapQ:45 readCov:13;
chr01build05r1.fasta samtools.snp 2044 2044 63 + . Note=G/T; Note=consQ:63 SNPQ:63 MaxMapQ:35 readCov:12;
chr01build05r1.fasta samtools.snp 2909 2909 51 + . Note=G/A; Note=consQ:51 SNPQ:51 MaxMapQ:29 readCov:8;
chr01build05r1.fasta samtools.snp 2965 2965 75 + . Note=T/G; Note=consQ:75 SNPQ:75 MaxMapQ:43 readCov:16;
chr01build05r1.fasta samtools.snp 4035 4035 69 + . Note=T/C; Note=consQ:69 SNPQ:69 MaxMapQ:33 readCov:14;
```

本日は DDBJ Pipeline の基礎部分を紹介



- バクテリアのNGSデータをアセンブル
- ヒトやマウスの公開されているトランスクリプトームをヒトゲノムの特定のバージョンの特定の染色体にマッピング
- などなど、ご興味のあるNGS公開配列をさがして、トライしてみてください
- もちろん、ご自身で取得したNGSデータをアップロードして解析することも可能です

【実習】基礎処理の流れをつかもう

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

DDBJ Read Annotation Pipeline

English Japanese

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。

LOGIN

新規アカウント作成 ゲストとしてログイン

ゲストとして
ログイン

Pipelineフローチャート

```
graph LR; User[ユーザ] -- "Reads, metadata" --> DRA[DDBJ Read Archive]; User -- "File Upload" --> FileUpload[Reads, metadata]; DRA -- "map positions" --> Foundation[基礎処理部]; Foundation -- "Mapping, de novo assembly" --> Archive[DDBJ]; Archive -- "metadata, annotation" --> Manual[Manual curation]; Manual -- "WGS/CON" --> Archive; Foundation -- "解析目的別ワークフロー 開発中" --> Workflows[ケノム解析, RNA-seq, ChIP-seq];
```

User ID:

Password:

動作中JOBの確認

PipelineのIDをお持ちでない場合、ゲストとしてログインすることができます。

マニュアルおよびチュートリアル

- 日本語チュートリアル
- 英語マニュアル
- DBCLS 統合TV チュートリアル1 - 今日からはじめるDDBJ Read Annotation Pipeline
- DBCLS 統合TV チュートリアル2 - DDBJ Read Annotation Pipelineによるde novo Assembly解析

Tweets Follow @pipeline_info

pipeline 29 Jul

処理を使うNGSの配列ファイルの用意

DDJB
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.
DDJB Read Annotation Pipeline.
Development Team.

Select Query Files → Select Tools → 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 : 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

FTP で手元から
アップロード可能

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

Workflow

Genome (SNP/Short
Indel)

RNA-seq (Tag count)

ChIP-seq

Select Query Files

Running Status

Set GenomeSet → Set Map Options → Confirmation

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

JOB STATUS

step1.
Preprocessing

step1.
Mapping

step1.
de novo Assembly

step2-All status

DELETE NEXT

HELP

HELP

TUTORIAL

Contact Us.
DDBJ Read Annotation
Pipeline.
Development Team.

処理を使うNGSの配列ファイルの用意

The screenshot shows the DDBJ DNA Data Bank of Japan pipeline interface. The main title is "Selecting Query Files". A large speech bubble on the right contains the Japanese text "公開データをインポート可能" (Importation of public data is possible). The interface includes a navigation bar at the top with steps: Select Query Files → Select Tools → Set QuerySet → Run → Confirmation. Below this are buttons for Running Status and a Logout link.

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 Query Files

NEXT

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 unchecke : 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
done	DRA000000	2012-08-22 11:01:10.00

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

NEXT

Selecting Query Files

FTP upload Private DRA entry Import public DRA Preprocessing HTTP upload

Metadata of the DRA entry.

Select a metadata file

TYPE	ACCESSION	ALIAS	FILENAME	DL
Submission	DRA000001	DRA000001	DRA000001.submission.xml	
Sample	DRS000001	DRS000001	DRA000001.sample.xml	
Study	DRP000001	DRP000001	DRA000001.study.xml	
Experiment	DRX000001	DRX000001	DRA000001.experiment.xml	
Run	DRR000001	DRR000001	DRA000001.run.xml	

STUDY TITLE Whole genome sequencing of *Baillus subtilis* subsp. *natto* BEST195
STUDY TYPE Whole Genome Sequencing

Select your registered query files.

Queries with different Instrument models can't be selected together.

single paired all clear

No.	Experiment ACCESSION	Sample ACCESSION	Run ACCESSION	STRAIN	Run_date	Read #	Read length	Instrument model	Layout
1	DRX000001	DRS000001	DRR000001	strain BEST195	2008-09-13	9,977,388	36	ILLUMINA	paired

: from metadata : Counted from query file (Read length is calculated from the first entry.)

DELETE NEXT

納豆菌の
公開データが
インポート済

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.
Preprocessingstep1.
Mappingstep1.
de novo Assembly

step2-All status

HELP

HELP

TUTORIAL

Contact Us.
DDJB 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			
				Base space	Color space	Paired end	Depth	Coverage	Error rate		SNP	Indel	.gff	.bed
<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 200 bp

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 severely 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 type="radio"/>	BLAT	Single-end analysis only

velvet で
アセンブル
しましよう

BACK NEXT

配列とペア形式のセットを選択

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

Generating Query Sets from Query Read Files

Paired-end analysis
Layout of paired sequence: 5'-3' 3'-5'
5' 3' 3' 5'
Linker(1) Target Linker(2) Linker(3) Target Linker(4)

	Run ACCESSION	Read length	Quality Score
<input checked="" type="checkbox"/>	D3R000001 ->-<	36 bp	Read1 Read2

Set as Mate-Pair Set as Pair-End

QUERY SET

RESET BACK NEXT

配列のセットの形式を選んで次へ

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

Generating Query Sets from Query Read Files

RESET BACK NEXT

Paired-end analysis
Layout of paired sequence: 5'-3' 3'-5'
5' 3' 5'
Linker(1) Target Linker(2) Linker(3) Target Linker(4)
Run ACCESSION Read length Quality Score
Set as Mate-Pair Set as Pair-End

QUERY SET
Query set1
PairedOrientation RunAccession RunAlias RowLength QualityScore1 QualityScore2
paired DRR000001 DRR000001 36

RESET BACK NEXT

今日はクエリファイルを1つしか選択していないので、このステップに意味がないが、複数のファイルを選択していた場合、それらをすべて結合して実行するか、あるいは、別々に連続して実行するかをここで選択します。

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

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

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 Options → Run

Run Confirmation

Destination of mail
When the request is completed, the system sends an email to this address:
yn@nig.ac.jp

Assembly [velvet]

Query sets
Query set1

PairedOrientation	RunAccession	RunAlias	RowLength	QualityScore1	QualityScore2
paired	DRR000001	DRR000001	36		

Assembly commands
velvet

Set optional parameters of the paired-end analysis

Step1) Convert sequences
Shuffle the sequence.
perl shuffleSequences_fastq.pl query_1.fastq query_2.fastq shuffle_query_pe.fastq
Running velveth.
Velveth output_directory/ 23 -fastq -shortPaired shuffle_query_pe.fastq

Step2) Assembly
Velvetg output_directory/ -ins_length 300 -exp_cov auto

Step3) Set parameters of the CONFIG mapping tool

Step4) Create assembled sequences in FASTA file from pileupped reads to [submit WGS division of DDBJ](#).

Set filtered length for contigs
-lengthfilter pl.pileupFile 100 out_WGS.txt

連絡先入力後
実行可能

guest では
実行できません

結果の確認・ダウンロード例

処理状況は
こちらから

DDBJ DNA Data Bank of Japan

Select Query Files → Select Tools → Set QuerySet → Set Ass. Options → Confirmation → Running Status

ACCOUNT
login ID [guest]
Logout

ANALYSIS
Data setup
DRA Start
FTP upload
HTTP upload
DRA Import
Preprocessing Start

JOB STATUS
step1. Preprocessing
step1. Mapping
step1. de novo Assembly (Red Box)
step2. All status

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

Status - de novo Assembly

Mapping Job de novo Assembly Job Preprocessing Job

Order
Sort by : ID Descending Show Only Your Own Job Reload

ID	User ID	Submission accession	P/S	Status	Tool	Read #	Read length	Assembly detail	Mapping detail	Start time	End time	Elapsed time
4914	guest	DRA000001 DRR000001	P	complete	SOAPdenovo	9,977,388	36			2013-01-11 22:06:40	2013-01-11 22:13:40	00:06:59
4912	guest	-- test.txt	S	complete	SOAPdenovo	1	--			2013-01-11 18:01:34	2013-01-11 18:02:16	00:00:42
4911	guest	-- preprocessing	S	error	SOAPdenovo		--			2013-01-11 --	2013-01-11 --	
4909	--	-- HPS1	P	complete	ABySS	5,754,246	--			2013-01-11 12:14:34	2013-01-11 13:52:21	01:37:47
4908	guest	DRA000001 2008-09-12.BES	P	complete	Velvet	9,977,388	36			2013-01-11 12:13:53	2013-01-11 16:10:01	03:56:08
4907	--	-- HPS1	P	complete	ABySS	5,754,246	--			2013-01-11 12:03:54	2013-01-11 13:24:40	01:20:45
4900	--	-- Ion_mt20mergec	S	complete	ABySS	148,666	--			2013-01-10 15:32:13	2013-01-10 --	00:16:28

page 1 NEXT >

View (Red Box)

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.gz	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

次はMappingの例: SRA044892 をロード

Accession :

Organism :

Arabidopsis thaliana

StudyType :

RNASeq

CenterName :

Platform :

ILLUMINA

シロイヌナズナ
の発現データ

Studies							<<	<	1	/ 2	Page	>	>>
STUDY_ID	STUDY_TITLE	STUDY_TYPE	ORGANISM	BASES	SUBMITTED	CENTER_NAME							
SRP007763	High-resolution profiling of small RNAs in the Arabidopsis thaliana root	RNASeq	Arabidopsis thaliana	9.2G	2011-05-24	Duke University							
2 SRP007763	SRA044892 Genome-wide detection of context-sensitive alternative splicing in Arabidopsis roots	RNASeq	Arabidopsis thaliana	1.3G	2011-08-11	Institute of Plant and Microbial Biology, Academia							
3 SRP007845	SRA045009 Genome-wide detection of context-sensitive alternative splicing in Arabidopsis roots	RNASeq	Arabidopsis thaliana	2.3G	2011-08-18	Institute of Plant and Microbial Biology, Academia							
4 SRP008262	SRA045978 GSE32216: SKIP Is a Splicing Factor Linking Alternative Splicing and Circadian Clock in Arabidopsis	RNASeq	Arabidopsis thaliana	3.7G	2011-09-19	GEO							
5 SRP008348	SRA046111 GSE32318: 2-week-old Arabidopsis seedlings (Columbia ecotype)	RNASeq	Arabidopsis thaliana	3.4G	2011-09-23	GEO							
6 SRP008486	SRA046358 GSE32284: IBM1, a JmjC domain histone demethylase, is involved in the regulation of RNA-directed DNA methylation through epigenetic control of RDR2 and DCL3 expression in Arabidopsis.	RNASeq	Arabidopsis thaliana	3.6G	2011-09-28	GEO							
7 SRP008822	SRA046998 Unexpected diversity of chloroplast non-coding RNAs as revealed by deep sequencing of the Arabidopsis transcriptome	RNASeq	Arabidopsis thaliana	9.8G	2011-10-12	Salk-E							
8 SRP009136	SRA047499 Alternative splicing landscape in Arabidopsis	RNASeq	Arabidopsis thaliana	17.6G	2011-10-28	MUW							
9 SRP009340	SRA048085 GSE33713: Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos	RNASeq	Arabidopsis thaliana	7.7G	2011-11-15	GEO							
10 SRP009369	SRA048118 Mapping Gene Activity of Arabidopsis Root Hairs	RNASeq	Arabidopsis thaliana	22.7G	2011-11-17	Institute of Plant and Microbial Biology, Academia							
11 SRP009850	SRA048671 GSE34476: Analysis of the Arabidopsis shoot meristem transcriptome during floral transition identifies distinct	RNASeq	Arabidopsis thaliana	3.8G	2011-12-15	GEO							

シロイヌナズナの転写解析データをロード

DDBJ
DNA Data Bank of Japan

ACCOUNT
login ID [guest]
Logout

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

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](#)

でも今は押さないで！

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	DRA000583	2012-08-20 14:04:46.08

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

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 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.)

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

 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.)

DELETE

NEXT

Bowtie2 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

SELECTING TOOLS FOR BASIC ANALYSIS OF DDBJ ANNOTATION PIPELINE

Reference Genome Mapping

	Tool	Help	Version	Base space	Color space	Paired end	Depth	Coverage	Error rate	SNP	Indel	.gff	.bed	SAM	Comment
<input type="checkbox"/>	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	✓		✓		✓		✓				✓	

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

Bowtie2 (selected) 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 severely 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

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: ✓ TAIR8, ✓ TAIR9, TAIR10

all check

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 変更なればそのまま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

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 -d10000000 -f refgenome.fasta out.bam | bcftools view -bvcg - >

out.var.raw.bcf
bcftools view out.var.raw.bcf | vcfutils.pl varFilter -D10000 > out.var.fltr.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

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 Options → Submit

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 --unmapped out.unmapped

Step3) Convert the read alignment to .BAM format
samtools view -bS -o out.bam out.sam

Step4) Detect DNA polymorphism

連絡先いれたら
実行可能

guest では
実行できません

「RUN を押した」と思ってください

処理状況は
こちらから

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
Processing

JOB STATUS
step1.
Preprocessing
step1.
Mapping (Red Box)
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	User ID	Submission accession	P/S	Status	Tool	Read #	Read length	Genome size	Detail	Start time	End time	Elapsed time
4915	guest	SRA044892 control	S	running	Bowtie2	5,925,048	--	121 M	View	2013-01-11 23:23:03	—	
4913	guest	SRA049447 9870	S	complete	Bowtie2	14,278,727	--	4 M	View	2013-01-11 18:21:55	2013-01-11 19:26:40	01:04:44
4910	—	ERA000092 1_dpf_assay1_1 1_dpf_assay1_2 1_dpf_assay1_3 1_dpf_assay2_1 and more...	P	running	Bowtie2	375,421,197	--	1,379 M		2013-01-11 16:19:14	—	
4906	—	— A1_Unshu_Paire	P	running	bwa	110,759,316	--	299 M		2013-01-10 16:24:05	—	
4905	—	— Ion_mt20mergec	S	complete	Bowtie2	148,666	--	16,520		2013-01-10 15:58:39	2013-01-10 16:04:15	00:05:36
4904	—	— Ion_mt20mergec	S	complete	Bowtie2	148,666	--	16,831		2013-01-10 15:54:22	2013-01-10 16:00:46	00:06:24
4903	—	—	S	complete	Bowtie2	148,666	--	16,589		2013-01-10 00:05:36		

JOB STATUSstep1.
Preprocessingstep1.
Mappingstep1.
de novo Assembly

step2-All status

HELPHELP TUTORIAL

Contact Us.

DDBJ Read Annotation
Pipeline.

Development Team.

- [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

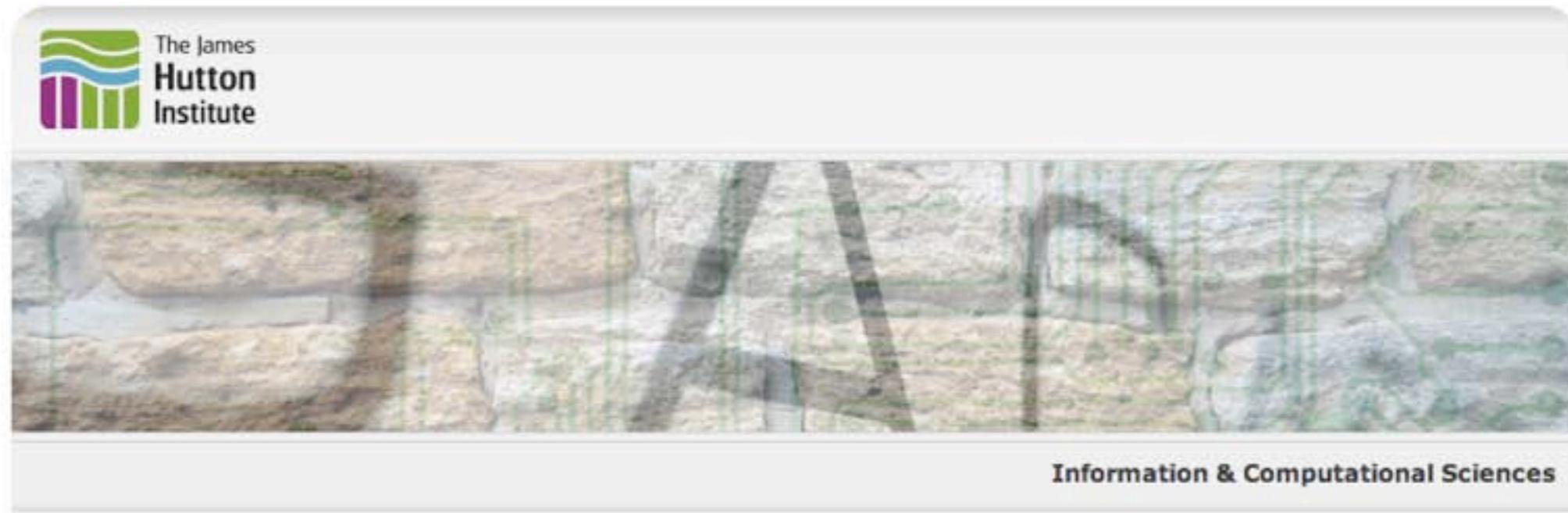
Wait time	Start time	End time
0: 1:12	2013-01-11 23:23:03	2013-01-12 00:19:10

実行結果

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
	samtools view -h out2.bam > out.sam	2013-01-11 23:45:13	2013-01-11 23:45:33			Download(334.4 MB)	MD5
	samtools pileup -c -f all.fa out2.bam > /home/w3pipeline/refdata/result/guest/4915/4729/17860/pileup/out.pileup	2013-01-11 23:47:05	2013-01-11 23:53:54		View	Download(400.6 MB)	MD5
	For DepthCoverage samtools pileup -c -f all.fa out2.bam > /home/w3pipeline/refdata/result/guest/4915/4729/17860/depth-coverage/out.pileup	2013-01-11 23:55:26	2013-01-12 00:02:15			Download(400.6 MB)	MD5
	samtools view -hX out2.bam > out.samX	2013-01-12 00:02:17	2013-01-12 00:04:00			Download(333.7 MB)	MD5

マッピング結果を眺めるには (BAM/SAM)

1. Tablet (<http://bioinf.scri.ac.uk/tablet/>)

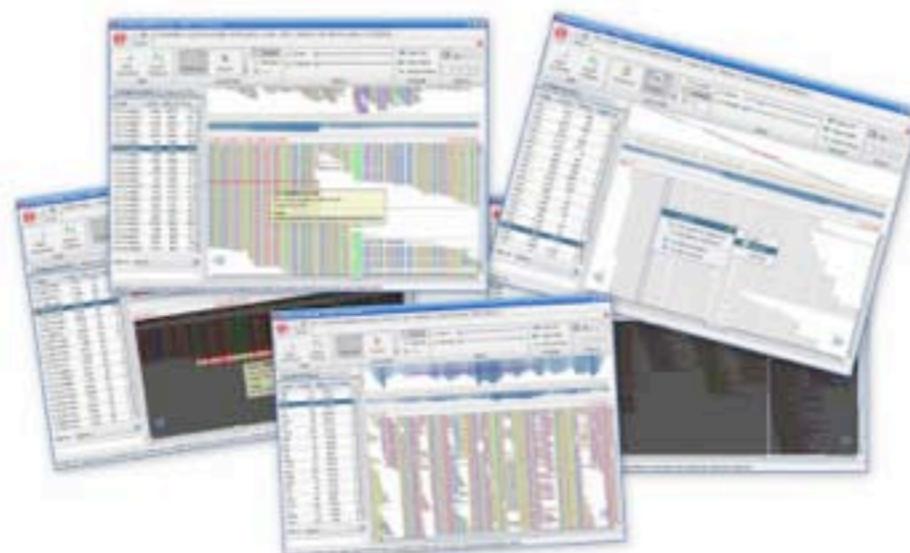


- Tablet**
- [Tablet Homepage](#)
 - [Download Tablet](#)
 - [Screenshots](#)
 - [Tablet FAQ](#)
 - [Sample Data](#)
 - [Assembly Conversion](#)
 - [Papers and Presentations](#)
 - [Privacy Policy](#)
 - [Tablet World Map](#)
 - [Online Help](#)

- Our Software**
- [CurlyWhirly](#)
 - [Flapjack](#)
 - [OPTIRas](#)
 - [Strudel](#)
 - [Tablet](#)
 - [TetraploidMap](#)
 - [TOPALI](#)

Tablet - Next Generation Sequence Assembly Visualization

Tablet is a lightweight, high-performance graphical viewer for next generation sequence assemblies and alignments.



マッピング結果を眺めるには (BAM/SAM)

2. IGV (<http://www.broadinstitute.org/igv/>)

The screenshot shows the homepage of the Integrative Genomics Viewer (IGV) at <http://www.broadinstitute.org/igv/>. The page features a large central image showing a screenshot of the IGV software interface, which displays genomic tracks and data. To the left of this image is the main navigation menu:

- Home
- Downloads
- Documents
 - Hosted Genomes
 - FAQ
 - IGV User Guide
 - File Formats
 - Release Notes
 - Credits
- Contact

Below the menu is a search bar labeled "Search website" with a "search" button. At the bottom left is the "BROAD INSTITUTE" logo and the text "© 2012 Broad Institute".

What's New

- NEWS** December 18, 2012. IGV 2.2 has been released. See the [release notes](#) for more details.
- April 20, 2012. IGV 2.1 has been released. See the [release notes](#) for more details.
- April 19, 2012. See our new [IGV paper](#) in *Briefings in Bioinformatics*.

Overview

The Integrative Genomics Viewer (IGV) is a high-performance visualization tool for interactive exploration of large, integrated genomic datasets. It supports a wide variety of data types, including array-based and next-generation sequence data, and genomic annotations.

Citing IGV

To cite your use of IGV in your publication:

Helga Thorvaldsdóttir, James T. Robinson, Jill P. Mesirov. [Integrative Genomics Viewer \(IGV\): high-performance genomics data visualization and exploration](#). *Briefings in Bioinformatics* 2012.

James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov. [Integrative Genomics Viewer](#). *Nature Biotechnology* 29, 24–26 (2011).

Funding

Development of IGV is made possible by funding from the National Cancer Institute, the National Institute of General Medical Sciences of the National Institutes of Health, and the Stowers Cancer Center.

New!

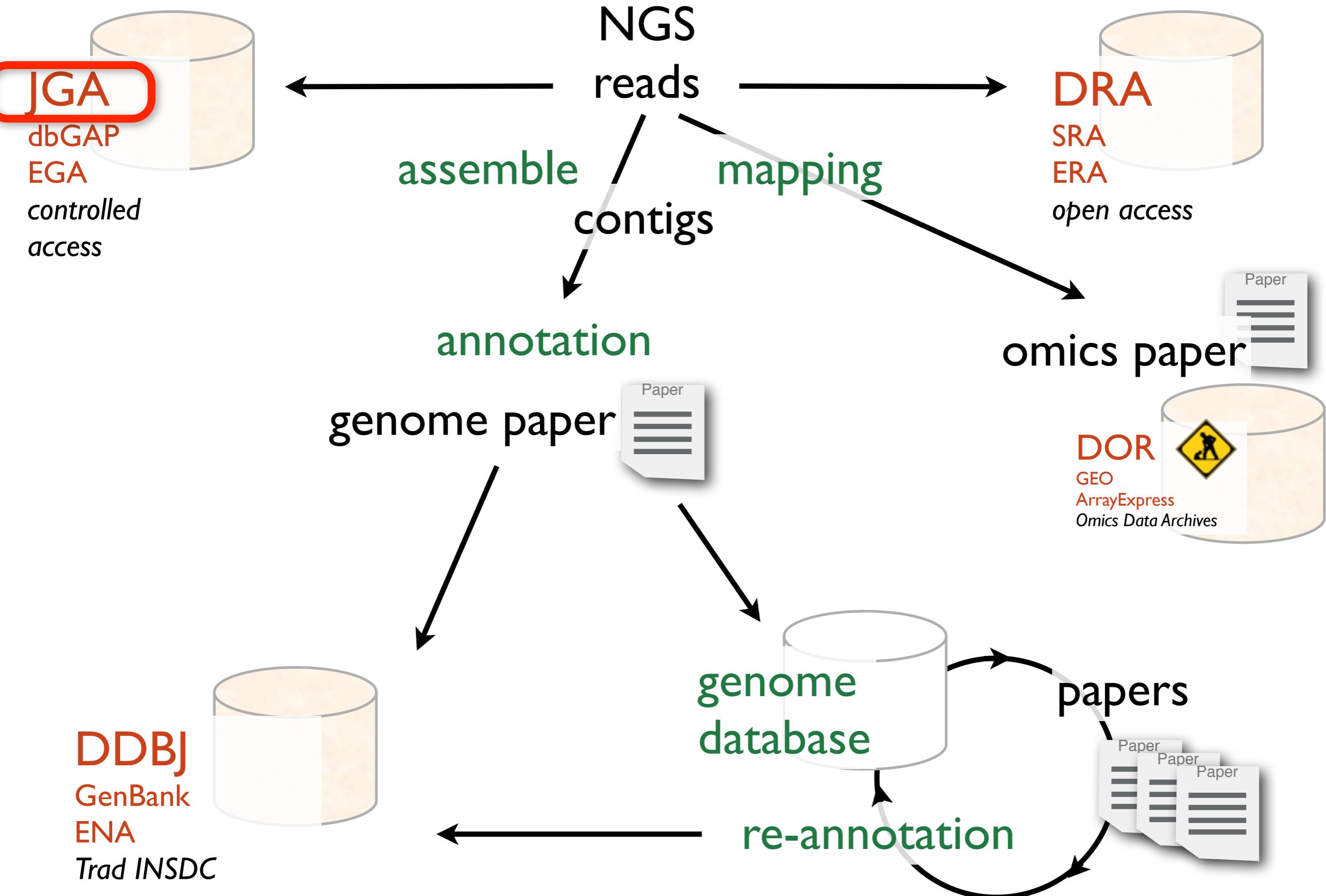
JGA

Japanese
Genotype-phenotype Archive

制限公開

NGSデータ

The ecosystem of sequence data



- ⌚ Japanese Genotype-phenotype Archive (JGA)
 - ⌚ 個人の遺伝学的なデータと表現型情報を保存し、提供
 - ⌚ 匿名化されているメタデータのみを受け付け
 - ⌚ Informed Consent により JGA データの利用は特定の研究目的に制限されている
 - ⌚ データの登録と利用は、事前に NBDC に申請し、認可されている必要がある
 - ⌚ メタデータの体系は SRA の拡張 (Dataset, Policy 追加)
 - ⌚ NGS データに加えアレイや variation データを登録できる
 - ⌚ ウェブサイト: <http://trace.ddbj.nig.ac.jp/jga/>
 - ⌚ メールアドレス: jga@ddbj.nig.ac.jp

 **DDBJ**
DNA Data Bank of Japan

Japanese Genotype-phenotype Archive

Home Studies Submission

概要

Japanese Genotype-phenotype Archive (JGA) は個人を特定される可能性のある遺伝学的なデータと表現型情報を保存し、提供しています。データが収集された個人との間の同意に基づく協定により、JGA のデータ利用は特定の研究目的に制限されています。JGA は厳格なプロトコールに従い、情報を管理、格納、提供しています。登録処理が終った全てのデータは暗号化されます。JGA チームにはこちらから連絡することができます。

利用制限ポリシー

JGA はデータに適用され
許可するかどうかは Nat
ます。利用者は NBDC
れた連絡を受け、利用者

データの登録

JGA は NBDC で認可された利用制限ポリシーを持つ、匿名化されたデータだけを受け付けています。登録者は NBDC に JGA へのデータ提供を申請し、JGA は NBDC から正式に提供が許可された連絡を受け、登録者に登録手順を案内します。多岐にわたるデータを受け付けており、アレイと新世代シーケンサーからの生データ、遺伝子型・構造変異、統計解析などの解析処理されたデータを保存しています。また、JGA はサンプルの表現型についてのデータを登録しています。

登録者はデータを JGA に投稿する前に JGA に連絡します。JGA チームが連絡してきた登録者に登録手順を別途案内します。

データは処理された後、投稿されたオリジナルなファイル形式でデータベースに登録されます。

データ登録/利用を申請

National Bioscience Database Center

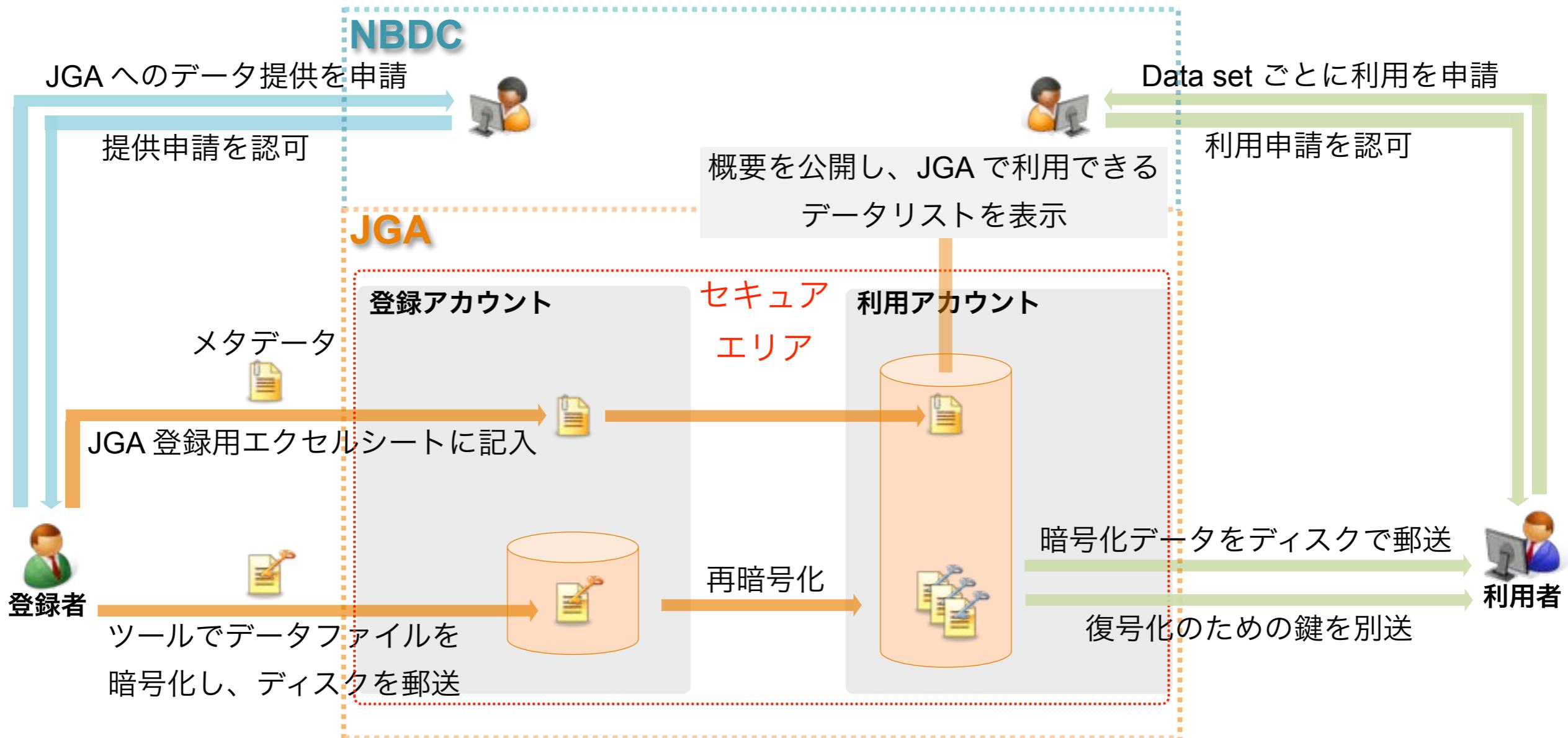
NBDC: 登録・利用を認可

DDBJ: アーカイブを担当

Sitemap Site Policy Privacy Contact © DNA Data Bank of Japan

Last modified: 2013-07-30

JGAの登録・利用フロー



当面、メタデータは Excel file で投稿し、配列データはディスク郵送で受けとる。今年度中に JGA 全体をオンライン化し、データの提供・利用をアカウント管理できるシステムに移行する。

(再掲) スパコン利用申請はこちら



- [遺伝研 スーパーコンピュータ] で検索

検索...

NIG SuperComputer

SuperComputer Facilities of National Institute of Genetics

現在地: Home

Language/言語

日本語

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)

システム情報

システム構成
システム使用方法
稼働スケジュール

各種申請・問い合わせ

各種申請

問い合わせ

重要なお知らせ

公開日	表題
2013年8月20日	計算機リソース拡張利用申請中止のお知らせ
2013年7月31日	バックアップサービス中止のお知らせ
2013年3月27日	NCBI ftpサイトへのアクセスについて
2013年2月4日	【再依頼】ディスク使用量削減のお願い

新着のお知らせ

公開日	表題
2013年8月26日	fatノードサービス再開のお知らせ
2013年8月23日	fatノードサービス障害のお知らせ
2013年8月22日	国立遺伝学研究所スーパーコンピュータシステム サービス復旧のお知らせ
2013年8月21日	fatノードサービス再開のお知らせ
2013年8月20日	fatノードサービス障害のお知らせ

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

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



DDBJ Read Annotation Pipeline

[English](#)[Japanese](#)

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。

LOGIN

新規アカウント作成 **ゲストとしてログイン**

User ID:
Password:
Login

動作中JOBの確認
PipelineのIDをお持ちでない場合、[ゲストとしてログインすることができます](#)。

マニュアルおよびチュートリアル

- [日本語チュートリアル](#)
- [英語マニュアル](#) ↗
- [DBCLS 統合TV チュートリアル1 - 今日からはじめるDDBJ Read Annotation Pipeline](#) ↗
- [DBCLS 統合TV チュートリアル2 - DDBJ Read Annotation Pipelineによるde novo Assembly解析](#) ↗

Tweets [Follow @pipeline_info](#)

pipeline 29 Jul

スパコンの利用申請後に

