

AJACS52@農工大

DDBJ/スパコン/ゲノム注釈

国立遺伝学研究所
大量遺伝情報研究室
中村保一



@yaskaz

a.k.a. catlover, ikasumipapa,
猫教授

使い倒し系バイオインフォマティスト



植物とか微生物のゲノム解析+DB屋



The Arabidopsis Genome Initiative
(2000) Analysis of the genome sequence
of the flowering plant *Arabidopsis thaliana*. *Nature*, 408, 796-815.

シロイヌナズナの 1/4
(27 Mb, 6200 genes) の解析



[http://genome.kazusa.or.jp/
cyanobase/](http://genome.kazusa.or.jp/cyanobase/)

光合成細菌のゲノム解析+データ
ベース。Social Bookmark による
遺伝子注釈系

1990年

京大院（典）

ゼニゴケミト

コンドリア

ゲノム

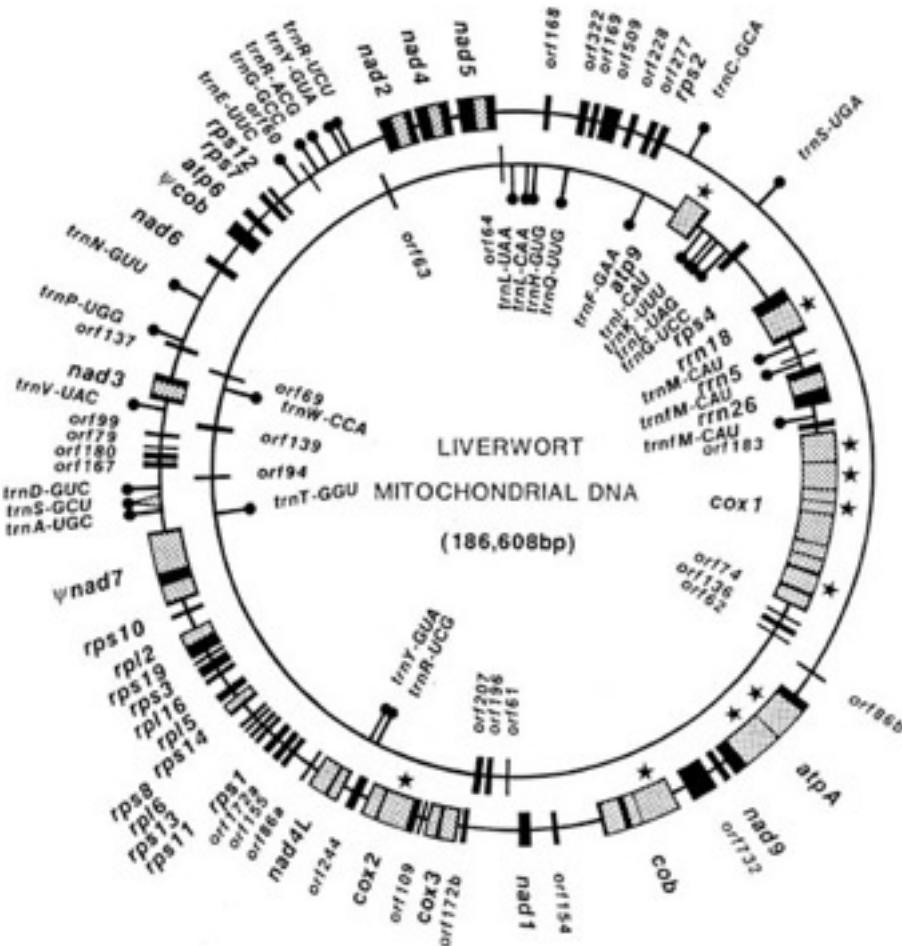
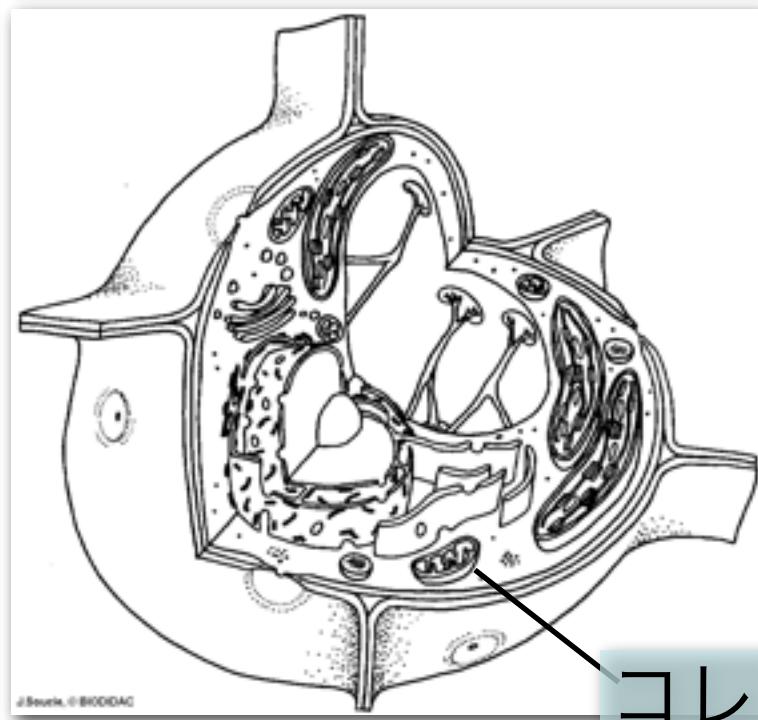


Fig. 1. Gene organization of the mitochondrial genome from a liverwort *Marchantia polymorpha*. Genes shown outside the map are transcribed anticlockwise, and those inside are transcribed clockwise. Solid and hatched boxes indicate exons and introns, respectively, in the coding regions. Asterisks indicate introns having ORF homologous to RNA maturase. Genes encoding tRNAs were localized by searching for T^Ψ loop sequence (GTTCRA) and then attempting to construct the typical clover-leaf structure, and identified by the one-letter amino acid code with their anticodons given in parentheses. The genes for the small subunit and large subunit ribosomal proteins are shown as *rps* and *rpl*, respectively. *atp*, *nad*, *cox*, and *cob* represent the genes for subunits of H⁺-ATPase, NADH-ubiquinone oxidoreductase, cytochrome c oxidase, apocytochrome b, respectively.

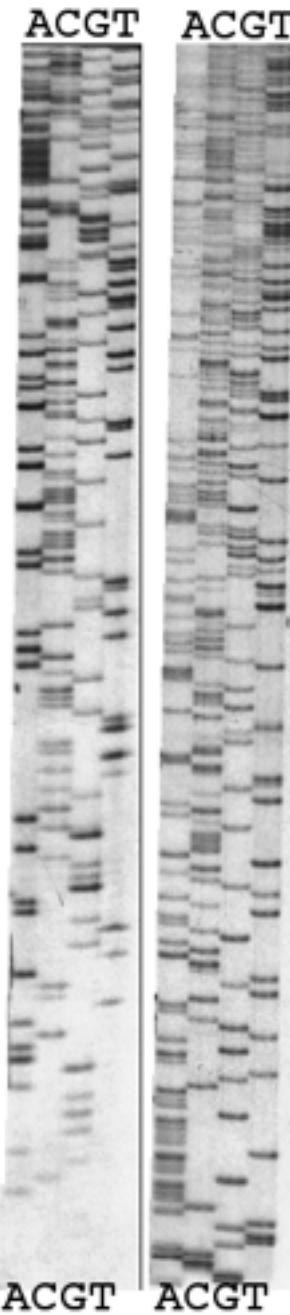


古典的配列決定 (dideoxy法) の原理

- Polymerase Chain Reaction
 - <http://www.youtube.com/watch?v=QaWLJVGEFi8>
- Sanger Sequencing (dideoxy method)
 - <http://www.youtube.com/watch?v=oYpllbbI0qF8>
 - <http://www.youtube.com/watch?v=6ldtdWjDwes>

【tips】

統合TVも便利だけど
他にもYouTubeで
実験系原理探すと
わかりやすいのがあるよ



イニシエの basecalling



digitizer

CCGCCCTAGTGGCGATGCCAGGGAAACAAACCGCTGCCGGGGCAATGGGTCGGACT
TGGCAGTCTCCTCCGGGGGATTGTACCTATCTCTGTGGTTAGCCACTGATTACCAATC
AGTTATGGTCCCCATCTAGTTCTATGGAGTGCTGGGCATTGCCAACGCGCTGATTCA
CACGGTATTCCAGTGCAAGTTAAATGGCCAATGACCTGTTATTGAAGGGAAAAAATTA
GCTGGCATTAAAACCGAAAGCAAATCAATGGAACAGAAATCACCGCCGCCATTGGG
GTGGGCATTAACTGGACTAACCCAGTACCAGCCACTGGCATTGCCCTAGGGCCTTTGT
GAAGCGGAATCAATCCAGAGTATCAACAGTCTGACGGATTAGCTGAAATTACCTGGCG
GGGCTCACCTGGTTGGCATCGTTACCAAAGAGAAGGCATTGCAGGCATTGGTAGAT
TATCTCCAATTATTCGCCATGGGGCCGGAAATTAGCCTAACCCAGGGAGTTGGCATA

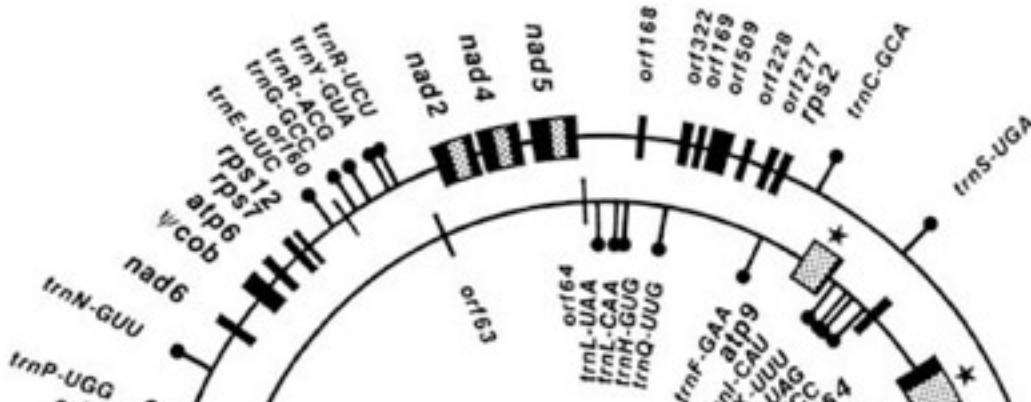


Communication

Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA [☆]: A primitive form of plant mitochondrial genome

Kenji Oda, Katsuyuki Yamato, Eiji Ohta, Yasukazu Nakamura, Miho Takemura, Naoko Nozato, Kinya Akashi, Takeshi Kanegae, Yutaka Ogura[†], Takayuki Kohchi[‡], Kanji Ohyama

186 kb



院生 7人

かかりで

まる 2年

- 1996 *Synechocystis* sp. PCC 6803
2000 *Arabidopsis thaliana*
2000 *Mesorhizobium loti*
2001 *Anabaena (Nostoc)* sp. PCC 7120
2002 *Bradyrhizobium japonicum*
2002 *Thermosynechococcus elongatus* BP-1
2003 *Gloeobacter violaceus* PCC 7421
2007 *Microcystis aeruginosa* NIES-843
2008 *Lotus japonicus*
2012 *Bradyrhizobium* sp. S23321
2012 *Solanum lycopersicum*
2012 *Eucalyptus globulus*
2012 *Hevea brasiliensis* (Para rubber tree)
2014 *Klebsormidium flaccidum* (an algae)
2014 *Weissella oryzae* SG25T
2014 *Lactobacillus oryzae* SG293T
2014 *Lactobacillus hokkaidonensis* LOOC260T
Marchantia polymorpha (a liverwort)
Citrus sinensis
Schizosaccharomyces japonicus

Cyanobacteria
Rhizobia
Plants
Others

Our on-going genome projects



a liverwort, *Marchantia polymorpha*
220 Mb genome
4.4 k scaffolds / N50: 1.3 Mb

a rubber tree, *Hevea brasiliensis*
1.4 Gb genome
52.7 k scaffolds / N50: 120 kb



a citrus tree, *Citrus unshiu*
360 Mb genome
21.1 k scaffolds / N50: 385 kb

遺伝研／DDBJは静岡県三島市にあります



DDBJ (<http://www.ddbj.nig.ac.jp/>)





English

Search

Google™ カス

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を読む

DDBJ Twitter

DDBJ Service

登録 Data Submission

検索・解析 Search / Analysis

スパコン Super Computer

アーカイブ ftp://ddbj.nig.ac.jp

Hot Topics

2013.06.26 WABI (Web API for Biology) の再開

2013.06.11 DDBJ リリース 93.0, DAD リリース 63.0 完成

2013.05.15 「第27回 DDBJing 講習会 in 三島(2013.7.4.開催)」のご案内 (参加申込み受付中)

Maintenance

Information

NIG 国立遺伝学研究所

大学共同利用機関法人 情報・システム研究機構

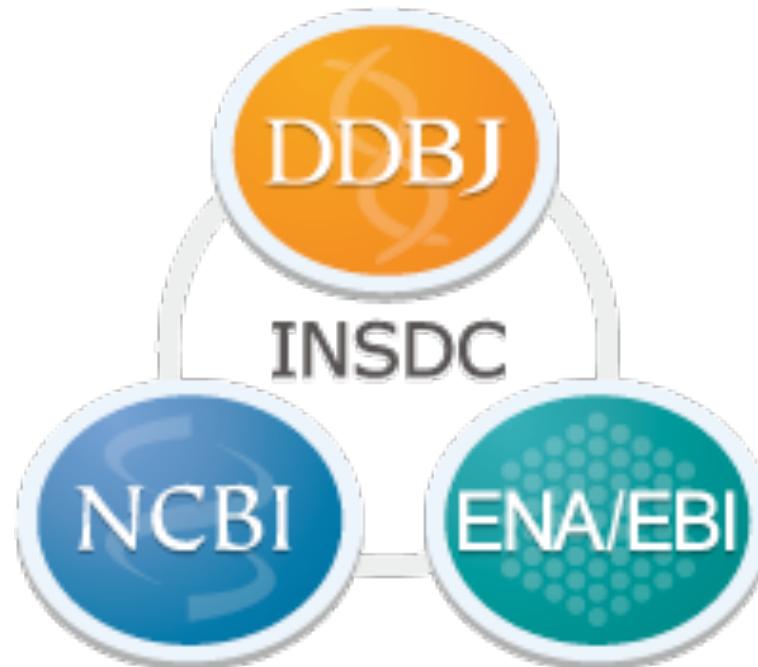
Research Organization of Information and Systems

塩基配列データバンクとはこのような事業

- 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



- 米国: GenBank (NCBI)
- 欧州: ENA (EBI)
- 日本: DDBJ



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

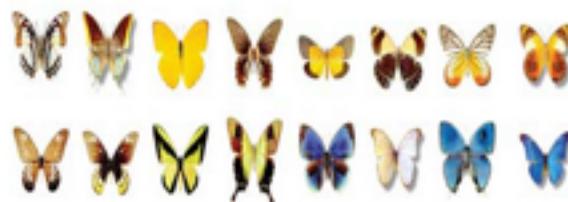
INSDCに多くの配列が登録された生物種

Images created by the Wordle.net web application are licensed under a Creative Commons Attribution 3.0 United States License.



DDBJに登録されている生物種 Top 100の ワードクラウド（数が多いほど大きい字で 表示）

NCBI Taxonomy (30万種)



Taxonomy

The Taxonomy Database is a curated classification and nomenclature for all of the organisms in the public sequence databases. This currently represents about 10% of the described species of life on the planet.

Using Taxonomy

[Quick Start Guide](#)

[FAQ](#)

[Handbook](#)

[Taxonomy FTP](#)

Taxonomy Tools

[Browser](#)

[Common Tree](#)

[Statistics](#)

[NameID Status](#)

[Genetic Codes](#)

[Linking to Taxonomy](#)

[Extinct Organisms](#)

Other Resources

[GenBank](#)

[LinkOut](#)

[E-Utilities](#)

[Batch Entrez](#)

[INSDC](#)

NCBI Taxonomy / *Felis catus*

NCBI Entrez PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books

Search for | as complete name lock Go Clear

Display 3 levels using filter: none

Felis catus

Taxonomy ID: 9685
Genbank common name: domestic cat
Inherited blast name: carnivores
Rank: species
Genetic code: [Translation table 1 \(Standard\)](#)
Mitochondrial genetic code: [Translation table 2 \(Vertebrate Mitochondrial\)](#)
Other names:

- synonym: *Felis silvestris catus*
- synonym: *Felis domesticus*
- common name: cats
- common name: cat
- includes: Korat cats
- authority: *Felis catus* Linnaeus, 1758

[Lineage \(full\)](#)

[cellular organisms](#); [Eukaryota](#); [Opisthokonta](#); [Metazoa](#); [Eumetazoa](#); [Bilateria](#); [Deuterostomia](#); [Chordata](#); [Craniata](#); [Vertebrata](#); [Gnathostomata](#); [Teleostomi](#); [Euteleostomi](#); [Sarcopterygii](#); [Dipnotetrapodomorpha](#); [Tetrapoda](#); [Amniota](#); [Mammalia](#); [Theria](#); [Eutheria](#); [Boreoeutheria](#); [Laurasiatheria](#); [Carnivora](#); [Feliformia](#); [Felidae](#); [Felinae](#); [Felis](#)

Entrez records	
Database name	Direct links
Nucleotide	53,444
Nucleotide EST	919
Nucleotide GSS	3,107
Protein	31,400
Structure	10
Genome	1
Popset	162
GEO Datasets	48
PubMed Central	892
Gene	25,772
SRA Experiments	83
Probe	2,889
Assembly	4
Bio Project	14
Bio Sample	72
Bio Systems	445
Clone DB	239,767
PubChem BioAssay	935
Protein Clusters	12
Taxonomy	1

English

Search

Google™カスタム検索

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を読む DDBJ Twitter

DDBJ INSDC NCBI ENA/EBI

International Nucleotide Sequence Database Collaboration

Hot Topics

Maintenance

Information

DDBJ Service

登録 Data Submission

検索・解析 Search / Analysis

スパコン Super Computer

アーカイブ ftp. ddbj.nig.ac.jp

2013.06.26 WABI (Web API for Biology) の再開

2013.06.11 DDBJ リリース 93.0, DAD リリース 63.0 完成

2013.05.15 「第27回 DDBJing 講習会 in 三島(2013.7.4.開催)」のご案内 (参加申込み受付中)

一覧

一覧

一覧

NIG 国立遺伝学研究所

大学共同利用機関法人 情報・システム研究機構

Research Organization of Information and Systems

生物学の 情報爆発

1000 Genomes

A Deep Catalog of Human Genetic Variation

[Home](#) [About](#) [Data](#) [Analysis](#) [Participants](#) [Contact](#) [Browser](#) [Wiki](#) [FTP search](#)

Search



LATEST ANNOUNCEMENTS

THURSDAY NOVEMBER 06, 2014

Phase3 variant calls for chrY are available, variant calls for chrX have been updated

Our final release of the Phase 3 variant set is now available on the [FTP site](#), including a newly added VCF file for chrY.

The chrY variant calls were made with a different process from that of the autosomes; a separate README is available in the release directory describing some details.

The chrX VCF file has been updated to include standard annotation including DP, continental super-population allele frequency.

The site file in the release directory is now wgs containing autosomes, chrX and Y.

Two algorithms were used to discover short tandem repeats (STRs) in the phase3 data. However the STRs did not make into the final integrated call set. They are now available separately [here](#).

The VCF files in the main release directory are also now available [here](#) in BCF format for faster processing time.

This release includes super population allele frequencies in the main release VCFs and [functional annotation](#) from the Ensembl Variant Effect Predictor along side many other datasets in the [supporting directory](#). The complete list of data is covered in the [Supporting Directory README](#). The issues which have been raised and resolved since our initial release are covered in the [Known Issues README](#).

Please send any questions about this data set to info@1000genomes.org

Recent project announcements

THURSDAY NOVEMBER 20, 2014

EMBL-EBI 1000 Genomes FTP site will be at reduced capacity between November 21th and December 8th

The EMBL-EBI FTP site will be at reduced capacity between November 21st and December 8th due to EMBL-EBI consolidating its web infrastructure into a single data centre.

NAVIGATION

- [Frequently Asked Questions](#)

LINKS

-  [All Project Announcements](#)
-  [Sample and Project Information](#)
-  [Media Archive](#)
-  [Download the 1000 Genomes Pilot Paper](#)

-  [Project Contacts](#)

-  [RSS Feed](#)

1001 Genomes

A Catalog of *Arabidopsis thaliana* Genetic Variation

[Home](#) [Data Providers](#) [Accessions](#) [Tools](#) [Software](#) [Data Center](#) [About](#) [Help desk](#)



Welcome to the 1001 Genomes Project

Check

Track the progress of genome sequencing and availability of *A. thaliana* accessions

[Go >](#)

Download

Use the Data Center to download project related SNPs, indels, SVs and genome sequences

[Go >](#)

Browse

Select, query and visualize polymorphisms of your favorite loci using POLYMORPH and GBrowse

[Go >](#)

Get Seeds

Seed sets of natural accessions are available for

80 strains (D. Weigel lab, MPI)
195 strains (J. Ecker lab, Salk)
180 strains (M. Nordborg Lab, GMI)

Links

- > GBrowse
- > WGS of 80 strains
- > Assemblies project
- > POLYMORPH
- > NCBI SRA Genomes Project
- > Map resource for 1001 Genomes

News

February 23, 2014
Chi-2 and Seattle-0 assemblies generated by SLU are now available in the [Data Center](#).

August 1, 2013
[SHOREmap v2.1](#) released.

August 6, 2012
Nd-1, sequenced by Center for Biotechnology of the University of Bielefeld (CeBiTec), is now available in the [Data Center](#). See the [project page](#) for more information.

November 29, 2011
JGI strains Bay-0 and Sha (both TAIR10) are now available in the [Data Center](#). See the [project page](#) for more information.

November 15, 2011
JGI strains (Alc-0, Blh-1, Jea, Oy-0, Ri-0 and Sakata) are now available in the [Data Center](#).

August 28, 2011
Cao et al. Whole-genome

The 1001 Genomes Vision

The 1001 Genomes Project was launched at the beginning of 2008 to discover the whole-genome sequence variation in 1001 strains (accessions) of the reference plant *Arabidopsis thaliana*. The resulting information is paving the way for a new era of genetics that identifies alleles underpinning phenotypic diversity across the entire genome and the entire species. Each of the accessions in the 1001 Genomes project is an inbred line with seeds that are freely available from the stock centre to all our colleagues. Unlimited numbers of plants with identical genotype can be grown and phenotyped for each accession, in as many environments as desired, and so the sequence information we collect can be used directly in association studies at biochemical, metabolic, physiological, morphological, and whole plant-fitness levels. The analyses enabled by this project will have broad implications for areas as diverse as evolutionary sciences, plant breeding and human genetics.

The complete genome sequences of over 80 accessions were released in early 2010 by the Max Planck Institute, and many more have been added since by the Salk Institute, the Gregor Mendel Institute and Monsanto. As of September 2014, over 1100 lines have been sequenced, and a publication that will describe an integrated analysis of the data is forthcoming.

- 特定の環境からサンプリングした生物相のDNAを、培養することなく全解析することができる
- “MetaGenomics”

Metagenomics is the study of metagenomes, genetic material recovered directly from environmental samples. The broad field may also be referred to as environmental genomics, ecogenomics or community genomics. (by Wikipedia)

現在進行中の配列決定プロジェクト

Home Search Distribution Graphs Biogeographical Metadata Statistics References Team Help News

Studies	20344
Biosamples	59030
Sequencing Projects	59165
Analysis Projects	41090

Download Excel Data file

Welcome to the Genomes OnLine Database

GOLD Release v.5

GOLD: Genomes Online Database, is a World Wide Web resource for comprehensive access to information regarding genome and metagenome sequencing projects, and their associated metadata, around the world.

Studies

- Metagenomic [546](#)
- Non-Metagenomic [19804](#)

Biosamples

- Classification
- Ecosystems
- Host-associated [11815](#)
- Engineered [1660](#)
- Environmental [6777](#)

Projects

- Complete Projects [6649](#)
- Permanent Drafts [23552](#)
- Incomplete Projects [26573](#)
- Targeted Projects [1404](#)

Organisms

- Organisms [60833](#)
- Archaea [1118](#)
- Bacteria [45965](#)
- Eukarya [9252](#)

Metagenome

Environmental

Register your genome or metagenome and Metadata in the Genomes Online Database

Register

Annotate your microbe or metagenome with IMG/ER or MER

Annotate

Publish your genome or metagenome in open access standards-supportive journal.

Publish

Please cite:

Reddy TBK, Thomas J, et al. The Genomes

Online Database (GOLD): A World Wide

Web Resource for Comparative Genomics.

(2014) doi: 10.1093/nar/gku950

<https://gold.jgi-psf.org/>

Lobos E and Kyripides N. The Genomes
Online Database: A World Wide Web
Resource for Comparative Genomics.
Nucl. Acids Res.

NGS

[次世代] Next-Generation Sequencer



[新型] New Generation Sequencer

新型シークンサーの特徴：高速・大量

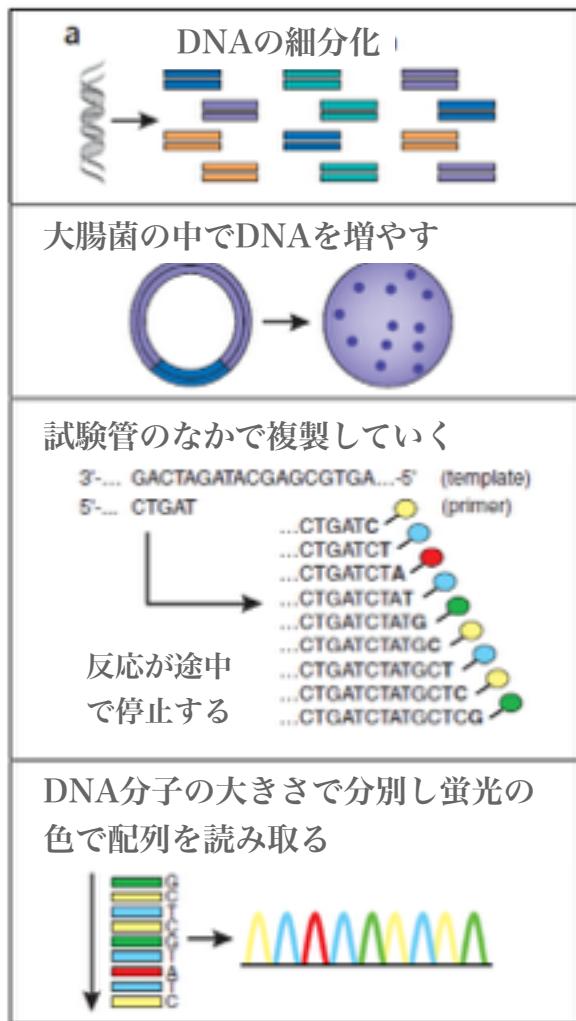


http://www.illuminakk.co.jp/systems/hiseq_systems.ilmn
より引用

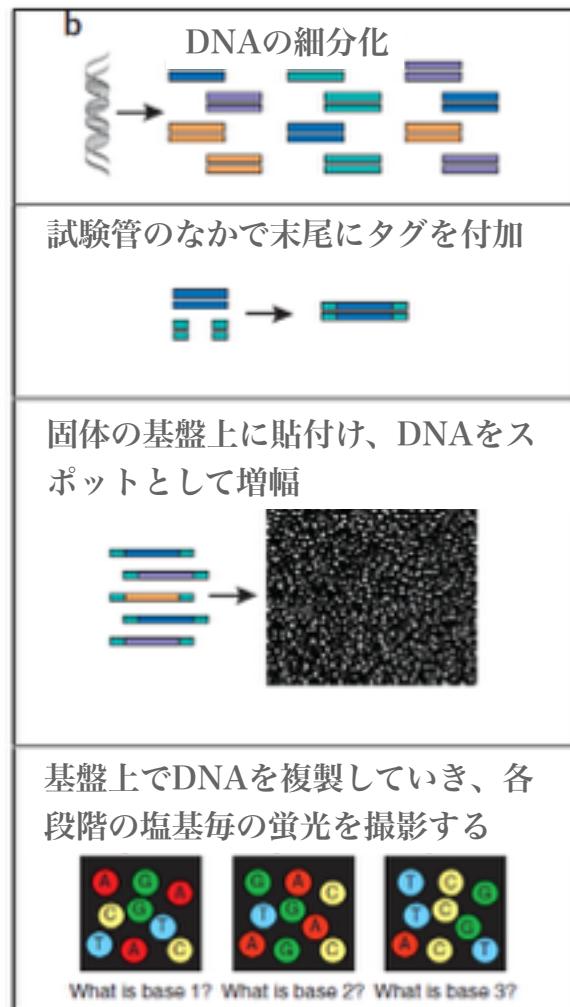
- ➊ イルミナ HiSeq 2500 / 2000
 - ➌ 一解析で6000億塩基 (600ギガベース)
 - ➌ ヒト一人のDNAがおよそ30億塩基対なので
 - ➌ 一解析でざっくり200人分ゲノムが取得できる

従来のシークエンサーと新型シークエンサー

従来法



新型



新型シーケンサはなぜ高速？→「集積度」

- ・ 従来法は溶液やゲル中での反応と分離
- ・ 固体担体を用いて超高密度化を可能にした



マイクロプレート
 $24 \times 16 = 384$ 穴



イルミナ社 GA フローセル
数千万スポット

さらに「ポータブル」シーケンサ

[search](#)[contact us](#)[for customers](#)[Home](#)[Technology](#)[About us](#)[News](#)[Careers](#)

Technology

Introduction to nanopore sensing

Electronics for nanopore sensing

The MinION™ device: a miniaturised sensing system

MinION™ Access Programme

A guide to MAP

The PromethION™ system

The GridION™ system

Workflow versatility: no fixed run time

Nanopore sensing: Informatics

Automatic optimisation of system performance

Analytes and applications: DNA, RNA, proteins

Fields of use

Publications

The MinION™ device: a miniaturised sensing system

Nanopore sensing technology can be miniaturised into a portable device for electronic single-molecule sensing.

The MinION™ device is a small instrument that is compatible with consumable flow cells containing the proprietary sensor chip, Application-Specific Integrated Circuit (ASIC) and nanopores that are needed to perform a complete single-molecule sensing experiment. Plugging directly into a laptop or desktop computer through a USB port, it is a self-contained device to deliver real-time experimental data.

The MinION device is adaptable for DNA sequencing, protein sensing and other nanopore sensing techniques. Currently, several hundred participants in the MinION Access programme (MAP) are working with the MinION system to explore how its features - including long read lengths, real time digital data, portability and compressed workflows - might help to answer a range of biological questions. The MAP initially focuses on DNA analysis.

The MinION is operated using MinKNOW™ software and participants in MAP will be performing basecalling in real time in the cloud, with the option to perform these analyses locally in the future.

Oxford Nanopore is focused on delivering the simplest possible sample preparation and workflows. The system is designed to be compatible with complex samples such as blood or serum and environmental samples such as water samples.

The MinION device is a miniaturised single-molecule analysis system, designed for single use and to work through the USB port of a laptop or desktop computer

The MinION device is designed to sense from complex samples such as blood/serum



DRA

DDBJ Sequence Read Archive

DDBJ Sequence Read Archive (DRA)

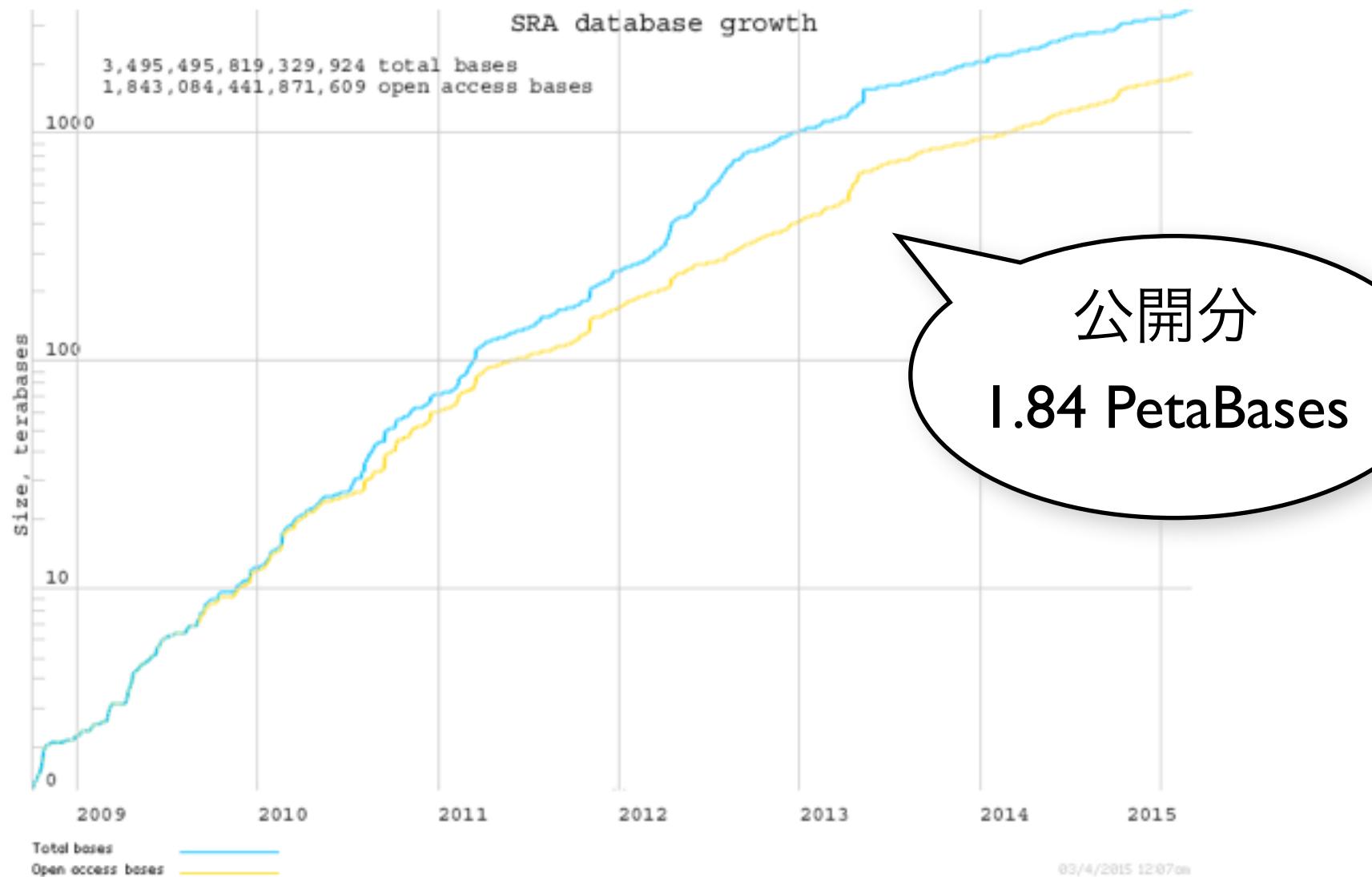


新世代シークエンサから出力される配列や
アライメントデータを登録・公開



SRA growth (NCBI)

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



DRA ウェブサイト ⇒ [DRA] で検索

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

登録関係情報

解析パイプライン

データ検索

データ取得

Home Submission ▾ Search Download ▾ Pipeline About

Login & Submit | Databases ▾ | English | Contact

Google™ カスタム検索

DDBJ DNA Data Bank of Japan

Sequence Read Archive

DDBJ Sequence Read Archive (DRA) は、Illumina 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD® System などの次世代シーケンサからの出力データを登録する国際協力組織である 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。従来のキャビラリーシーケンサからのデータ登録も可能で、DDBJ Sequence Read Archive にご登録ください。

検索

データをキーワード、生物名、シーケンサなどで検索する

登録

新型シーケンサからの生データやアライメントデータを登録する

動画マニュアル

DRA の利用方法や登録方法を解説している動画を見る

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

The screenshot shows the DRA Search interface. At the top, there is a search bar with 'Accession' set to 'DRP000001'. Below it are filters for 'Organism' (Pseudotrophus defective ATCC 49176), 'StudyType' (Epigenetics), and 'Platform' (ILLUMINA). A large black oval highlights the search bar area with the text '生物名 etc での絞り込み' (Filtering by organism name etc.). Below the search bar is a toolbar with 'Show 20 records', 'Sort by Study', 'Search', and 'Clear' buttons.

Statistics section on the left displays:

Type	Count
Submission	23770
Study	3423
Experiment	29624
Sample	111241
Run	71620

Organism section lists:

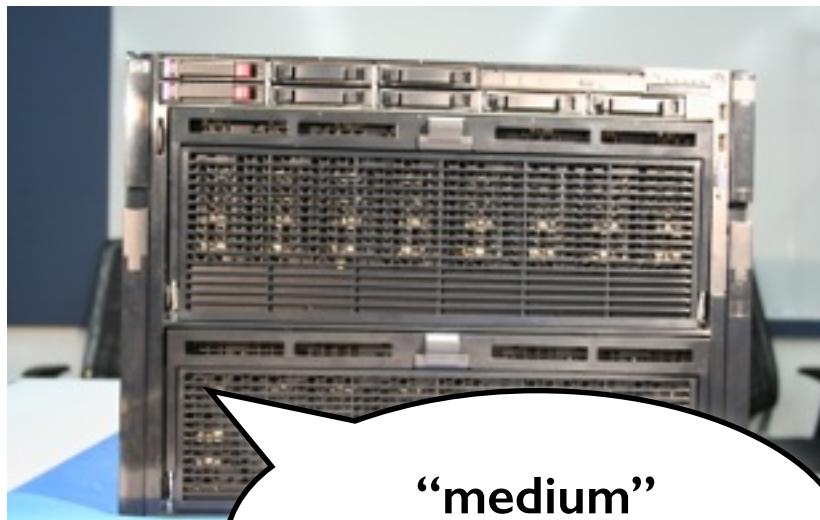
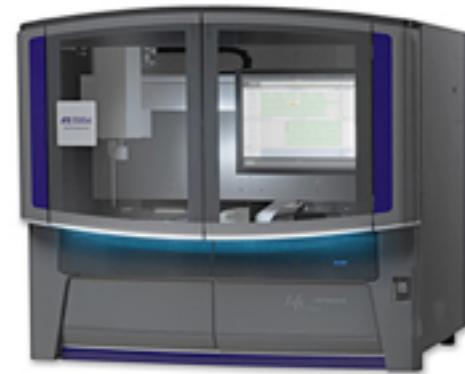
#	Organism Name	Study	#
1	Homo sapiens	293	1
2	metagenome sequence	169	2
3	Mus musculus	163	3
4	Drosophila melanogaster	121	4
5	murine metagenome	79	5
6	Ceaeorhabditis elegans	39	6
7	Arabidopsis thaliana	38	7
8	synthetic construct	37	8
9	Saccharomyces cerevisiae	35	9
10	Panicum virgatum	21	10

The main content area shows 'Search Results (358 studies)' with a table of studies. One study is selected: DRP000003. A blue arrow points from the search bar to this study detail page. The study detail page includes sections for 'Study Detail' (Title, Abstract, Description, Project ID, Center Name), 'Navigation' (Submission, Experiment, Sample), and a 'Download' button. A large black oval highlights the 'Download' button with the text 'ダウンロード' (Download).

A final black oval at the bottom right highlights the 'Description' section with the text '詳細 (メタデータ記述)' (Detailed description (metadata)).

詳細 (メタデータ記述)

NGS's + SC's in Biology



“medium”
2TB memory
x 10



“fat”
10TB memory
(SGI UV)

遺伝研スーパー

コンピュータ

English

Search

Google™カスタム検索

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を読む DDBJ Twitter

DDBJ INSDC NCBI ENA/EBI

International Nucleotide Sequence Database Collaboration

Hot Topics

2013.06.26 WABI (Web API for Biology) の再開
2013.06.11 DDBJ リリース 93.0, DAD リリース 63.0 完成
2013.05.15 「第27回 DDBJing 講習会 in 三島(2013.7.4.開催)」のご案内 (参加申込み受付中)

Maintenance

Information



DDBJ

DNA Data Bank of Japan

DDBJ Service

登録 Data Submission

検索・解析 Search / Analysis

スパコン Super Computer

アーカイブ ftp://ddbj.nig.ac.jp

Hot Topics

2013.06.26 WABI (Web API for Biology) の再開
2013.06.11 DDBJ リリース 93.0, DAD リリース 63.0 完成
2013.05.15 「第27回 DDBJing 講習会 in 三島(2013.7.4.開催)」のご案内 (参加申込み受付中)

Maintenance

Information

遺伝研スーパーコンピュータサイト top



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

The screenshot shows the homepage of the NIG Supercomputer Facilities of National Institute of Genetics. The header includes the NIG logo, the text '大学共同利用法人 情報・システム研究機構 国立遺伝学研究所 スーパーコンピュータシステム SuperComputer Facilities of National Institute of Genetics', and links for 'サイトポリシー' and 'サイトマップ'. The main content area features a '重要なお知らせ' (Important Notices) section with a list of announcements, a summary of the system, and images of server racks.

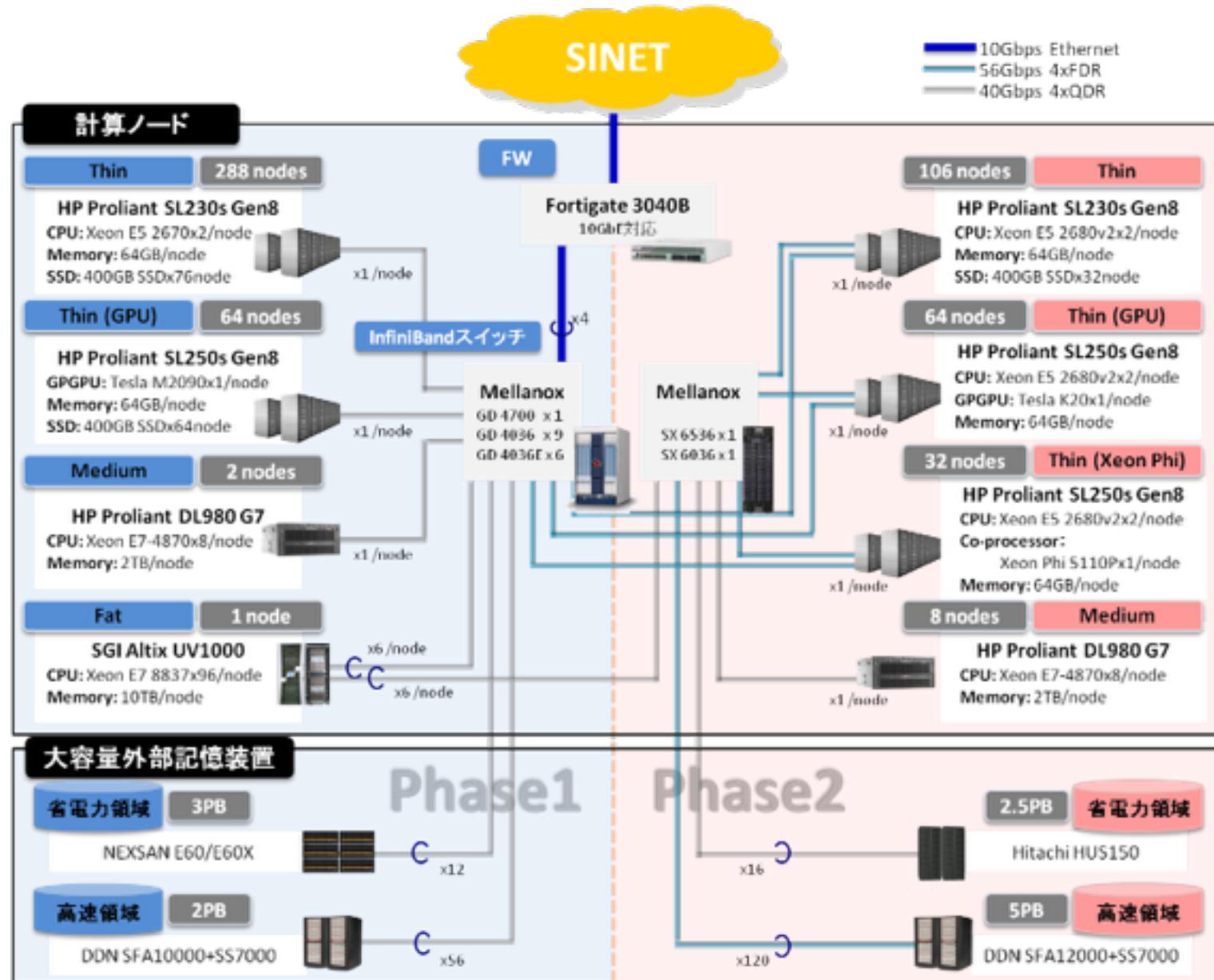
重要なお知らせ

公開日	表題
2014年7月22日	【スパコンユーザ会】三島開催 会場変更のお知らせ
2014年7月18日	ソフトウェインストールのお知らせ
2014年7月15日	【fat障害：2】fat計算ノードサービス再開のお知らせ
2014年7月14日	【通信障害：2】Phase1システム通信不具合の復旧のお知らせ
2014年7月14日	【通信障害】Phase1システム通信不具合のお知らせ
2014年3月4日	2014年3月8日からのスパコンPhase2システムご利用方法について

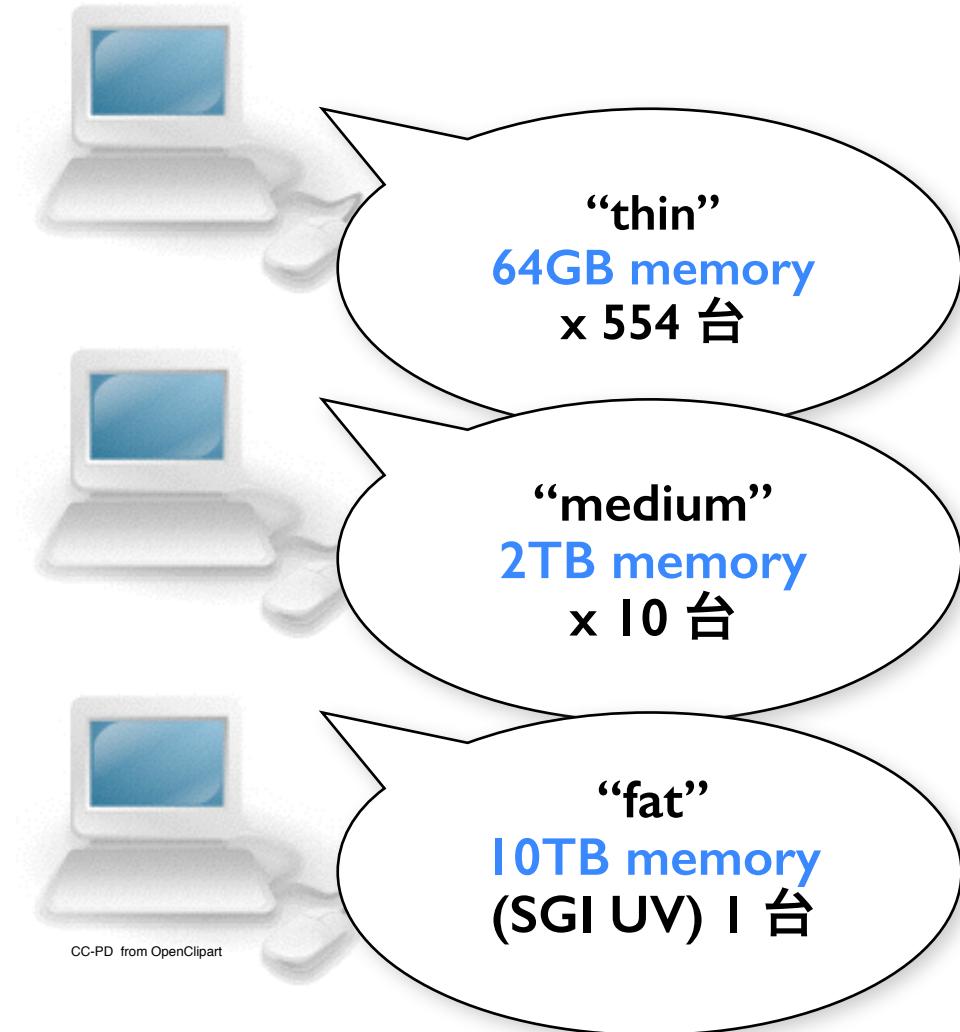
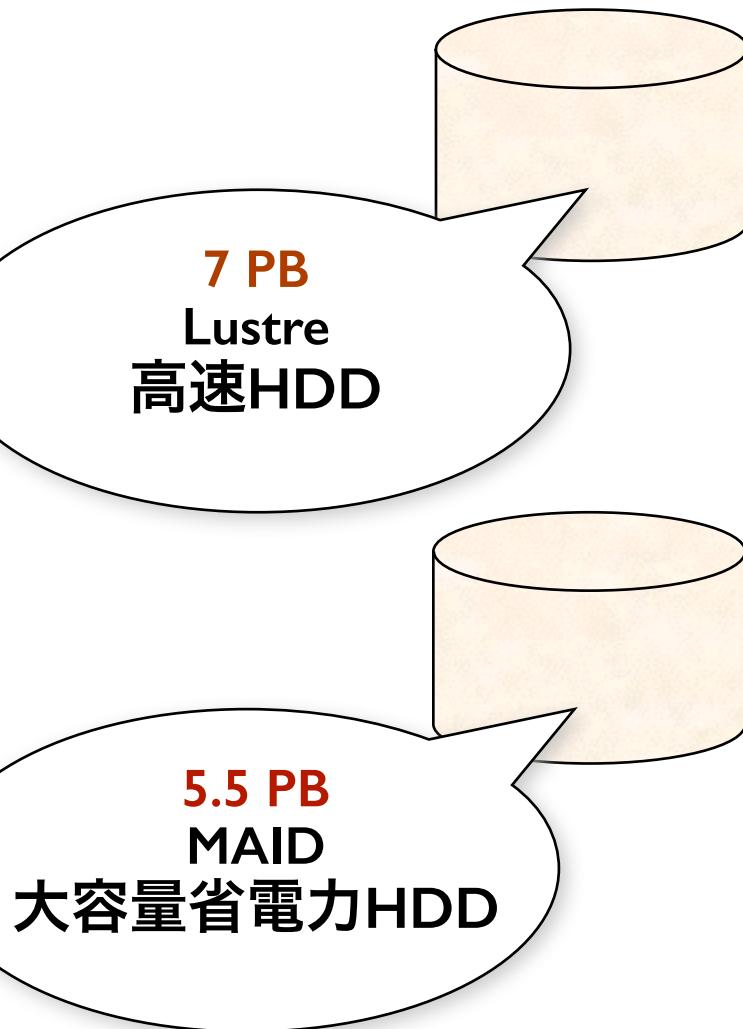
国立遺伝学研究所 スーパーコンピュータシステム(NIG SUPERCOMPUTER)とは

大学共同利用機関法人 情報システム研究機構 国立遺伝学研究所は、2012年3月にスーパーコンピュータシステムを更新しました。新しいスーパーコンピュータシステムはゲノム解析を主な目的とした大規模計算機利用拠点として最新鋭の大規模クラスタ型計算機、大規模メモリ共有型計算機、および大容量高速ディスク装置で構成されたスーパーコンピューティングシステムサービスを提供しています。

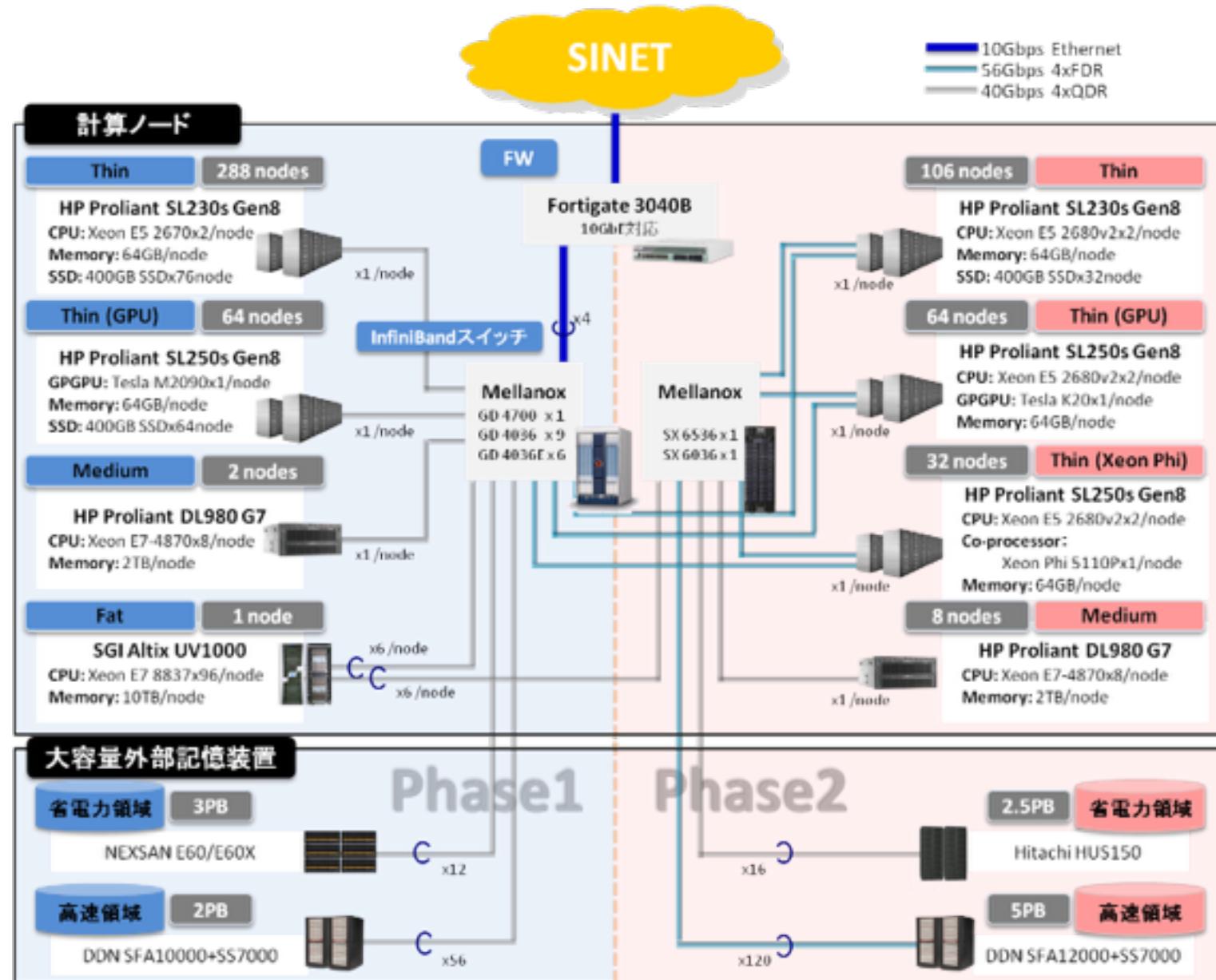
遺伝研スーパーコンピュータ (全容)



遺伝研スーパーコンピュータ (概要)



遺伝研スーパーコンピュータ (全容)



解析パイプラインも提供しています

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

Screenshot of the DDBJ Sequence Read Archive (DRA) website:

The top navigation bar includes: DDBJ logo, Login & Submit | Databases | English | Contact, Google™ カスタム検索, and a search icon.

The main menu bar includes: Home, Submission, Search, Download, Pipeline (highlighted with an orange arrow and yellow box), About, and a Pipeline link (also highlighted with an orange arrow and yellow box).

The Pipeline section contains the following information:

DDBJ Sequence Read Archive (DRA) は Roche 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD® System などの次世代シーケンサからの出力データのためのデータベースです。DRA は International Nucleotide Sequence Database Collaboration (INSDC) のメンバーであり、NCBI Sequence Read Archive (SRA) と EBI Sequence Read Archive (ERA) との国際協力のもと、運営されています。従来のキャ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。

Three buttons in the Pipeline section:

- 検索**: 検索アイコン (magnifying glass).
データをキーワード、生物名、シーケンサなどで検索する
- 登録**: 登録アイコン (ドライブ).
新型シーケンサからの生データやアライメントデータを登録する
- 動画マニュアル**: 動画アイコン (camera).
DRA の利用方法や登録方法を解説している動画を見る

DRA pipeline: ソフトウェア

よく用いられる
解析用ソフトウェアを
用意。クリックだけで
実行可能



ACCOUNT

login ID [yaskaz]

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**

Running Status

Selecting Tools for Basic Analysis of DDBJ ANNOTATION PIPELINE

① Reference Genome Mapping

			Input data			Evaluation			Analysis		Output format				
Tool	Help	Version	Base space	Color space	Paired end	Depth	Coverage	Error rate	SNP	Indel	.gff	.bed	SAM	Comment	
BLAT		34	✓					✓						Single-end analysis only	
Maq		0.7.1	✓		✓			✓	✓	✓	✓	✓	✓		
bwa		0.5.9	✓		✓			✓					✓		
SOAP		2.21	✓		✓			✓	✓	✓			✓		
Bowtie		0.12.7	✓	✓	✓			✓	✓				✓		
TopHat		1.0.11	✓		✓			✓					✓		
Bowtie2		2.0.0	✓	✓	✓			✓	✓				✓	For reads longer than about 50 bp, Bowtie2 is generally faster, more 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.
TIGAL		r2013-02-					

DRA pipeline: 比較対象

The screenshot shows the DDBJ DRA pipeline interface. At the top, there are three blue buttons: "Select Query Files", "Select Tools", and "Set QuerySet". Below them are two smaller buttons: "Running Status" and "Help".
ACCOUNT
login ID [yaskaz]
Logout
Change password
ANALYSIS
Data setup
DRA Start
FTP upload
Arabidopsis thaliana
Oryza sativa japonica
Oryza sativa indica
Zea mays B73
Sorghum bicolor
Homo sapiens
Mus musculus
Pan troglodytes
Caenorhabditis elegans
Xenopus (Silurana) tropicalis
Oryzias latipes
Solanum lycopersicum Heintz 1706
Saccharomyces cerevisiae
ChIP-seq
JOB STATUS
step1.
Preprocessing
step1.
Mapping
step1.
de novo Assembly
step2-All status
HELP

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 version 5.0
- tigr version 6.0
- tigr version 6.1
- 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 パイプライン、体験してみましょう

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



DDBJ Read Annotation Pipeline

English

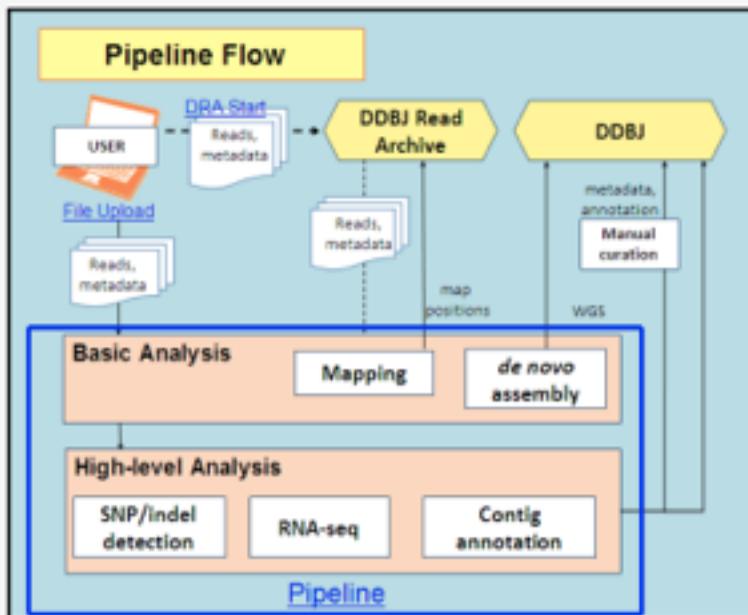
Japanese

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

L O G I N

New account

Login as "guest"



User ID: [Input field]

Check current jobs

- by the guest account.

ゲストとして
ログイン

Manual & tutorial

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

Tweets

Follow

pipeline

11 Jun

処理に使うNGSの配列ファイルの用意

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 → 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

The screenshot shows the DDBJ analysis pipeline interface. On the left, there's a sidebar with sections for ACCOUNT (login ID [guest], Logout), ANALYSIS (Data setup, DRA Start, FTP upload, HTTP upload, DRA Import, Preprocessing Start, step-1, Preprocessing, Mapping / de novo Assembly, step-2, Workflow (Genome (SNP/Short Indel), RNA-seq (Tag count), ChIP-seq)), JOB STATUS (step1. Preprocessing, step1. Mapping, step1. de novo Assembly, step2-All status), and HELP (HELP, TUTORIAL, Contact Us). The main area shows the 'Select Query Files' step with a progress bar. Below it, a table lists uploaded files via FTP:

	Filename	Description	Layout	Instrument model	File size
<input type="checkbox"/>	GSM727564_c0Foxh1.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

Buttons for DELETE and NEXT are located at the bottom right of the file list.

処理に使うNGSの配列ファイルの用意

The screenshot shows the DDBJ Pipeline interface. The top navigation bar includes 'Select Query Files', 'Select Tools', 'Set QuerySet', 'Run', and 'Confirmation'. Below this is a 'Running Status' section. The main content area is titled 'Selecting Query Files' and features a 'NEXT' button. A large speech bubble highlights the 'Import public DRA' tab, which is highlighted in yellow. Below it, a section titled 'Import public FASTQ files from DRA database.' contains instructions and a form for inputting a DRA accession number. A second form to the right provides a link to 'DRA Search'. At the bottom, there's a table of job status with columns for 'Status', 'Submission', and 'Request date'.

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.
DOBJ Read Annotation
Pipeline.
Development Team.

Select Query Files → Select Tools → Set QuerySet → Run → 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

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.

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	DRA000086	2012-08-24 13:51:17.364

今回はupload済のエントリから

DDBJ
DNA Data Bank of Japan

ACCOUNT
login ID [guest]
Logout

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

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

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

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

Running Status

NEXT

Selecting Query Files

FTP upload Private DRA entry Import public DRA Preprocessing HTTP upload

Metadata of the DRA entry.

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

Select a metadata file

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

ACCOUNT

login ID [guest]

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

Reference Genome Mapping

Tool	Help	Version	Input data			Evaluation			Analysis		Output format			Comment
			Base space	Color space	Paired end	Depth	Coverage	Error rate	SNP	Indel	.gff	.bed	SAM	
BLAT		34	✓					✓						Single-end analysis only
Maq		0.7.1	✓			✓			✓	✓	✓	✓	✓	
bwa		0.5.9	✓			✓			✓					
SOAP		2.21	✓			✓			✓	✓	✓			
Bowtie		0.12.7	✓	✓	✓				✓	✓				
TopHat		1.0.11	✓			✓			✓					
Bowtie2		2.0.0	✓	✓	✓				✓	✓				For reads longer than about 100 bp

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.
<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.
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
BLAT	Single-end analysis only

velvet で
アセンブル
しましよう

配列のセットの形式を選んで次へ

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
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.

Generating Query Sets from Query Read Files

RESET BACK NEXT

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

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

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 Analysis → 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 velvet.

Velvet 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
 and length filter of alignment >= 100
 and length filter of alignment <= 1000

連絡先いれたら
実行可能

でも今は
押さないで！

「RUN を押した」と思ってください

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
step1. de novo Assembly

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

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

Status - de novo Assembly

Mapping Job de novo Assembly Job Preprocessing Job

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

Delete * page 1 NEXT >

ID	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		00:06:59
4912	guest	---	S	complete	SOAPdenovo	1	—			2013-01-11 18:01:34		00:00:42
4911	guest	---	S	error	SOAPdenovo	—	—			2013-01-11 18:02:16		—
4909	—	HPS1	P	complete	ABYSS	5,754,246	—			2013-01-11 12:14:34		01:37:47
4908	guest	DRA000001 2008-09-12.BES	P	complete	Velvet	9,977,388	36			2013-01-11 12:13:53		03:56:08
4907	—	HPS1	P	complete	ABYSS	5,754,246	—			2013-01-11 12:03:54		01:20:45
4900	—	Ion_mt20mergec	S	complete	ABYSS	148,666	—			2013-01-10 15:32:13		00:16:28

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.
CDDBJ Read Annotation
Pipeline.
Development Team.

Detail view

BACK

Job info

ID

4914

Tool (Version)

SOAPdenovo (1.05)

RunAccession or Filename

DRR000001

Download

[DRR000001.fasta.gz](#)
[DRR000001_1.fasta.gz](#)
[DRR000001_2.fasta.bz2](#)

Read length

36 bp

Alias

DRR000001

Download modified queries

- [DRR000001_1.fasta.gz](#) (Original size 1.7 GB)
- [DRR000001_2.fasta.gz](#) (Original size 1.7 GB)

Download wgs file

- [out_WGS.fasta.gz](#) (Original size 3.9 MB)

Assembly statistics

Contig # : 5,300
 Total contig size : 4,138,179
 Maximum contig size : 49,938
 Minimum contig size : 24
 N50 contig size : 13,255

アセンブル結果の
基本情報

Time

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

Command	Start time	End time	Log1	Log2	Result	MD5
SOAPdenovo127mer all -s soapdenovo.conf -o output	2013-01-11 22:06:40	2013-01-11 22:12:28	View	View	Download(174.7 MB)	MD5

結果ファイル

BACK

Mappingの例 (DRAsearch+pipeline)

シロイヌナズナ
alternative splicing

Send Feedback ▶ Search Home ▶ DRA Home

StudyType : Platform :

Keyword : alternative splicing Arabidopsis

Show 20 records Sort by Study Search Clear

Search Results (2263 records) << < 1 / 114 > >>

Filtered by

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

#	META_FILE	ACCESSION	STUDY	STUDY_TITLE	STUDY_TYPE	ORGANISM	BASES	SUBMITTED	CENTER_NAME
1	SRA050132.study.xml .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	SRA044892.study.xml of context-sensitive alternative splicing in Arabidopsis roots </STUDY_TITLE> <STUDY_TYPE>existing	SRP007763	SRP007763	Genome-wide detection of context-sensitive alternative splicing in Arabidopsis roots	Other	Arabidopsis thaliana	1.3G	2011-08-11	Institute of Plant and Microbial Biology, Academia

データのIDはこちら

SRP007763

Study Detail	
Title	Genome-wide detection of context-sensitive alternative splicing in Arabidopsis roots
Study Type	Other
Abstract	To analyze context-sensitive changes in pre-mRNA splicing pattern and gene expression, we mapped the transcriptome of iron-deficient and iron-sufficient Arabidopsis roots using the RNA-seq technology. RNA-seq data were analyzed with a newly developed software package, RACKJ (Read Analysis & Comparis .. [more]
Description	
Center Name	Institute of Plant and Microbial Biology, Academia

Navigation

 Submission	SRA044892	 FTP
 Experiment	SRX102000	 FASTQ  SRA
 Sample	SRS256250	

p.ddbj.nig.ac.jp を開き、さっきのIDを入力

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
DRA Import

NEXT

Selecting Query Files

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

Import public FASTQ files from DRA database.

Here is do the section of automatic download of public DRA/ERA/SRA entries.
Please input DRA/ERA/SRA accession number. Then the pipeline system import metadata and FASTQ files from DRA database.

Input DRA/ERA/SRA Accession Number
SRA044892 Add my DRA entry

Accession Number can find here.
[DRA Search](#)

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

NEXT

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.
Preprocessingstep1.
Mappingstep1.
de novo Assembly

step2-All status

HELP

HELP

TUTORIAL

 Contact Us:
DDBJ Read Annotation
Pipeline,
Development Team.

Selecting Query Files

NEXT

FTP upload

Private DRA entry

Import public DRA

Preprocessing

HTTP upload

Metadata of the DRA entry.

Select a metadata : SRA044892

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

STUDY TITLE

Genome-wide detection of context-sensitive alternative splicing in Arabidopsis roots

STUDY TYPE

RNASeq

Select your registered query files.

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

No.	Experiment ACCESSION	Sample ACCESSION	Run ACCESSION	STRAIN	Run_date	Read #	Read length	Instrument model	Layout
1	SRX092046	SRS256250	SRR331219					ILLUMINA	single
2	SRX092046	SRS256250	SRR331224					ILLUMINA	single

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

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

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**

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
Maq	0.7.1		✓		✓			✓	✓	✓	✓	✓	
bwa	0.5.9		✓		✓			✓				✓	
SOAP	2.21		✓		✓			✓	✓	✓		✓	
Bowtie	0.12.7		✓	✓	✓			✓	✓			✓	
TopHat	1.0.11		✓		✓			✓				✓	

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

配列を選んで confirm, 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.

Generating Query Sets from Query Read Files

RESET BACK NEXT

Single analysis
Layout of single sequence:
5' Linker(1) Target Linker(2) 3'

Read Number	Read Length	Quality Score
<input checked="" type="checkbox"/> SRR331219 >	bp	

confirm

RESET BACK NEXT

TAIR10（最新）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.

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 変更なければそのまま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 (Red circle)

bowtie2
Set optional parameters of the single-end analysis

Step1) Convert reference sequence
bowtie2-build -f refgenome.fasta bt2-idx

Step2) Map
bowtie2 -q -p 4 -x bt2-idx -U query1.fastq(fasta) -S out.sam -u out.unmapped

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

Step4) Detect DNA polymorphism
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 | vcftools pl varFilter -D10000 > out.var.flt.vcf

Step5) Analysis for Depth, Coverage
samtools sort -o out.bam out_sorted.bam
samtools pileup -c -f reference.fa out_sorted.bam > out.pileup
perl pileup_for_CoverageDepth.pl out.pileup reference.fa
* This command does not appear in the list.

Step6) Create assembled sequences in FASTA file from pileupped reads to submit WGS division of DDBJ.

終了したらメールが来ます

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 → Step 4
Running Status

Run Confirmation

Destination of mail
When the request is completed, the system sends an email to this address:

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.fasta(-fasta) -S out.sam --u
out.unmapped

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

Step4) Detect DNA polymorphism

連絡先いれたら
実行可能

でも今は
押さないで！

BACK RUN

「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
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	01:04:44	<div style="width: 100%; background-color: green; height: 10px;"></div>
4910	—	ERA000092 1_dpl_assay1_1 1_dpl_assay1_2 1_dpl_assay1_3 1_dpl_assay2_1 and more...	P	running	Bowtie2	375,421,197	—	1,379 M	View	2013-01-11 16:19:14	—	
4906	—	— A1_Unshu_Pair	P	running	bwa	110,759,316	—	299 M	View	2013-01-10 16:24:05	—	
4905	—	— Ion_mt20mergec	S	complete	Bowtie2	148,666	—	16,520	View	2013-01-10 15:58:39	00:05:36	<div style="width: 10%; background-color: green; height: 10px;"></div>
4904	—	— Ion_mt20mergec	S	complete	Bowtie2	148,666	—	16,831	View	2013-01-10 15:54:22	00:06:24	<div style="width: 10%; background-color: green; height: 10px;"></div>
4903	—	—	S	complete	Bowtie2	148,666	—	16,589	View	2013-01-10 16:00:46	00:05:36	<div style="width: 10%; background-color: green; height: 10px;"></div>

DNA Import	Tool (Version)
Preprocessing Start	Bowtie2 (2.0.0)
step-1	
Preprocessing	RunAccession or Filename
Mapping / de novo Assembly	Download
step-2	SRR331219
Workflow	SRR331219.fastq.bz2
Genome (SNP/Short Indel)	Read length
RNA-seq (Tag count)	N.A. bp
ChIP-seq	Alias
JOB STATUS	
step1.	
Preprocessing	
step1.	
Mapping	
step1.	
de novo Assembly	
step2-All status	
Position errors	
Map ratio	
PDF download	
total query # : 5,925,048	
mapped query # : 5,037,456	
map ratio : 85.020 %	
Depth, Coverage	
coverage : 29886366 / 119482012 * 100 = 25.013	
depth : 384468141 / 29886366 = 12.864	

実行結果

Development Team.

Time

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

all.fa

Command	Start time	End time	Log1	Log2	Result	MD5
bowtie2-build -f all.fa refgenome	2013-01-11 23:23:03	2013-01-11 23:25:26	View		Download(194.4 MB)	MD5
bowtie2 -p 4 -q -x refgenome -U SRR331219.fastq -S out.map --un out.unmapped	2013-01-11 23:26:07	2013-01-11 23:38:02		View	Download(463.0 MB)	MD5
samtools view -bS -o out.bam out.map	2013-01-11 23:40:56	2013-01-11 23:42:28		View	Download(488.9 MB)	MD5
samtools sort out.bam out2	2013-01-11 23:42:59	2013-01-11 23:44:52		View	Download(357.5 MB)	MD5

- 「バイオインフォマティクス人材養成カリキュラム NGS 速習コース」で検索
- 企画・スケジュール
 - <http://events.biosciencedbc.jp/training/ajacs47>
- 動画・資料
 - http://biosciencedbc.jp/human/human-resources/workshop#NGS_sokusyu_2014
- 少々長いのですが（二週間のコース）突っ込んで勉強するには、こちらも役立つと思います