

# 統合データベース講習会：AJACS出島

## 2014年7月3日

# ゲノムデータベース、 次世代シーケンスデータベースの使い方

---

情報・システム研究機構（ROIS） ライフサイエンス統合データベースセンター（DBCLS）  
科学技術振興機構（JST） バイオサイエンスデータベースセンター（NBDC）

河野 信

# 本日の資料

## ◆ MotDB (Master of the DataBase)のページに あります <http://motdb.dbcls.jp/>

- ”motdb”を検索エンジンで検索 → AJACS出島の「講習会資料」をクリック

### MotDB

#### MotDB

ようこそMotDB（「もっとでいーびー」と発音、Master of the database【日本語訳：データベースの達人】の略）へ。このサイトでは、ライフサイエンス分野のデータベース統合を担う人材AJACS(All Japan Annotator/Curator/System DB administrator)向けの教材を提供しています。このサイトは、文部科学省委託研究開発事業「統合データベースプロジェクト」（平成18年度～平成22年度）の元で運用されてきましたが、引き続き、ライフサイエンス統合データベースセンター(DBCLS: Database Center for Life Science)により維持管理されています。

- 統合データベース講習会 AJACS
  - 募集！あなたの地域で「統合データベース講習会」を開催してみませんか？
    - 今年度の募集は終了しました
  - 今後の予定
    - 統合データベース講習会 : AJACS出島 (2014年7月3日開催・長崎大学) [講習会資料](#)
    - 統合データベース講習会 : AJACS信濃 (2014年7月17日開催・信州大学) [講習会資料](#)
    - 統合データベース講習会 : AJACS阿波 (2014年8月20日開催・徳島大学)
    - 統合データベース講習会 : AJACS十勝 (2014年9月5日開催・帯広畜産大学)
    - 統合データベース講習会 : AJACS岩手 (2014年12月5日開催・岩手医科大学)



統合データベース講習会  
**AJACS 出島**  
統合データベース講習会は、生命科学系のデータベースやツールの使い方。  
データベースを統合する活動を紹介する講習会です。

● 日程: 2014年7月3日(木) 9:20~17:45

● 会場: 長崎大学ボンバー会館1階セミナー室

● 対象: 生命科学分野のデータベースを利用したい、研究に役立てたい方。

● 内容: 今回の講習会は、生命科学系データベースのカタログ、種別表、アーリーアクセス情報、データベースの組合せ、データベースの組合せ等、各種に次世代シーケンストラックの実用的な解説例について紹介します。  
参加者全員がノンプログラミングでデータベースを使いながらの講習会。

● 定員: 約10名 ● 料金: 無料

● 応募方法: 下記ページよりお申込みください。  
<http://events.bioinfo.nibc.ac.jp/jaacs/spec/>  
(締切: 2014年6月11日(土) 17:00 締め切り 4月5日(木)まで)

● PCご自分でPCをお持ちの方のみOK。

● プログラム  
9:00～9:20 「受け渡し」  
9:20～10:30 「「ゲムデータベース」次世代シーケンストラックの使い方」 ……河野 佳 (DBCLS)  
10:30～11:00 休憩  
11:00～12:30 「「ゲムデータベース」次世代シーケンストラックの使い方」 ……真鍋 実理 (NEDO)  
12:30～13:30 「「化合物データベース」、「スライドデータベース」の使い方」 ……村松 啓明 (DBCLS)  
13:30～14:45 休憩  
14:45～15:45 「次世代シーケンストラックの使い方」 ……真鍋 実理 (NEDO)  
15:45～16:45 「次世代シーケンストラックの使い方」 ……村松 啓明 (DBCLS)  
16:45～17:45 「次世代シーケンストラックの使い方」 ……真鍋 実理 (NEDO)  
17:45～18:00 「次世代シーケンストラックの使い方」 ……村松 啓明 (DBCLS)  
主催: 科学技術振興機構「バイオインフォマティクスセンター(NEDO)」  
共催: 情報システム研究機構「バイオインフォマティクスセンター(NEDO)」、キムラー・カントン・システムズ  
統合データベース講習会は、生命科学系のデータベースやツールの使い方。  
データベースを統合する活動を紹介する講習会です。

統合データベース講習会  
**AJACS 信濃**  
統合データベース講習会は、生命科学系のデータベースやツールの使い方。  
データベースを統合する活動を紹介する講習会です。

# 本日の資料

- ◆ 「新しいタブで開く」と元の画面を残したまま、新しい画面を開くことができます
  - ・右クリック、もしくはコントロールを押しながらクリック

## プログラム

- 09:20～10:50 「ゲノムデータベース、次世代シーケンスデータベースの使い方」 / 河野 信（情報・システム研究機構 ライフサイエンス統合データベースセンター／科学技術振興機構 バイオサイエンスデータベースセンター）
- 10:50～11:00 休憩
- 11:00～12:30 「NBDCとNBDCが提供するサービス」 / 箕輪 真理（情報・システム研究機構 ライフサイエンス統合データベースセンター／科学技術振興機構 バイオサイエンスデータベースセンター） 講義資料 [横断検索（講習会用）](#)
- 12:30～13:30 昼食
- 13:30～15:30 「化合物データベース・パスウェイデータベースの紹介」 / 時松 敏（データベースセンター）
- 15:30～15:45 休憩
- 15:45～17:45 「次世代シーケンスデータの視覚化」 / 佐藤 哲也（九州大学 生体防

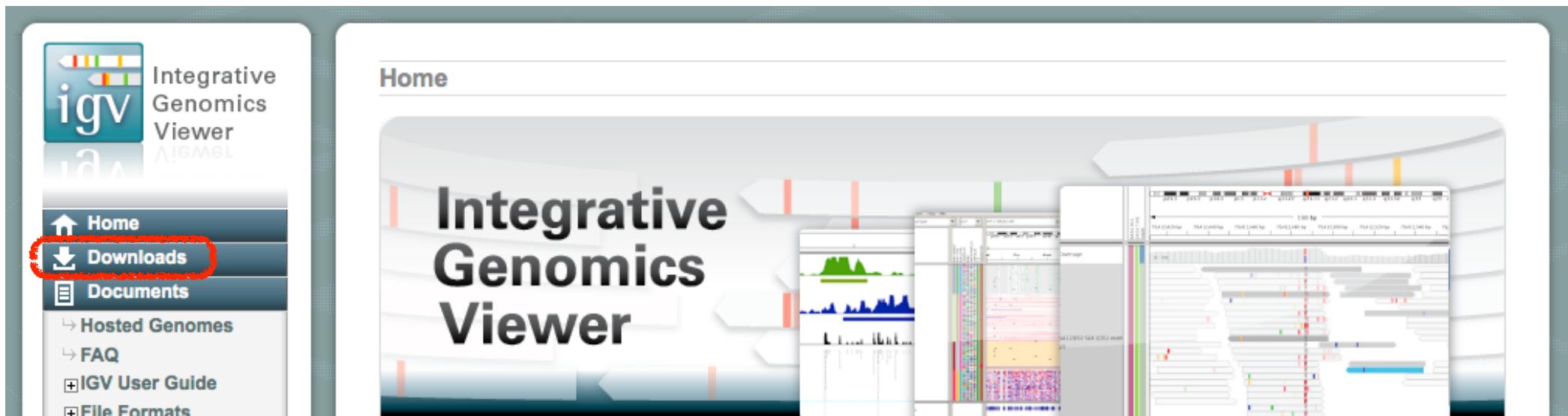


## お問い合わせ

統合データベース講習会事務局 AJACS at [biosciencedbc.jp](http://biosciencedbc.jp) までメールにてお問い合わせください

# IGVのインストール

- ◆ 午後の「次世代シーケンスデータの視覚化」でIntegrative Genomics Viewer (IGV)を使用
- ◆ あらかじめダウンロードをお願いします
  - <http://www.broadinstitute.org/software/igv/home>
  - “IGV” もしくは “IGV genomics” で検索
  - ユーザ登録が必要



# 本日紹介する内容（ゲノムデータベース）

- ◆ ゲノムデータベースの概要
- ◆ 【実習】 Ensembl
  - ゲノムを表示するためのブラウザ
- ◆ 【実習】 BioMart
  - ゲノム情報を取得するためのウェブインターフェイス
- ◆ 【実習】 GGRNA/GGGenome
  - 超高速遺伝子検索/ゲノム配列検索

# 本日紹介する内容（次世代シーケンスDB）

---

- ◆ 次世代シーケンスデータベースの概要
- ◆ 【実習】 DRAsearch
  - ◆ 次世代シーケンスデータベース検索
- ◆ 【実習】 DDBJ解析パイプライン
  - ◆ ウェブ上で次世代シーケンスデータを解析する
- ◆ Galaxy/MiGAP
  - ◆ 配列データのアノテーション

# DNA塩基配列データベース

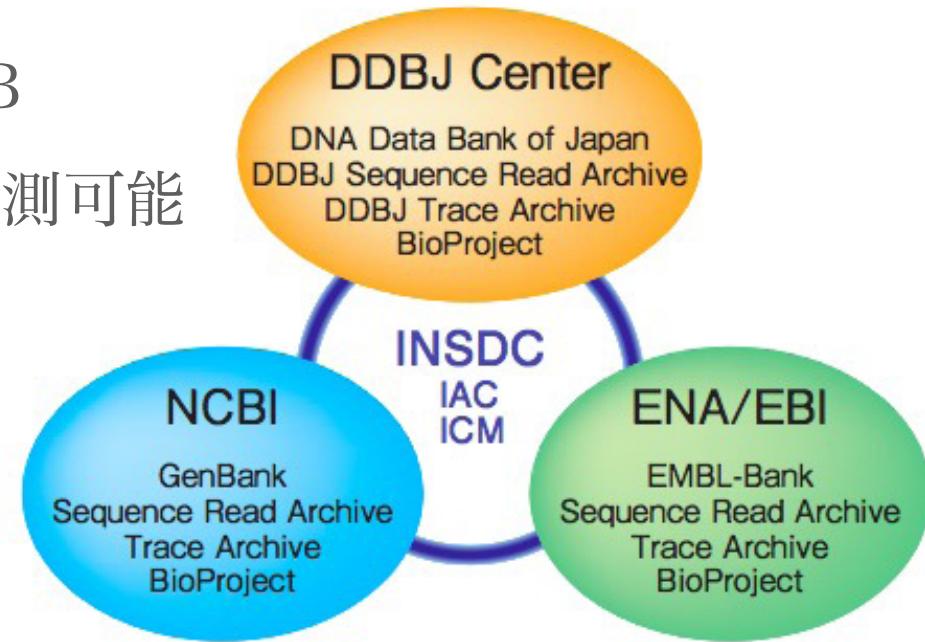
## ◆ GenBank (NCBI)/ENA (EBI)/DDBJ (NIG)

- DNA塩基配列を決定したらどちらかに登録する
- 3極でデータを毎日交換しており、データの内容は同じ

## ◆ Refseq

- NCBIがGenBankを再整理したDB
- IDによってどのようなデータか推測可能

AC\_ 複数個体から集められた完全ゲノム配列  
NC\_ 1個体から集められた完全ゲノム配列  
NG\_ 不完全なゲノム配列  
NM\_ mRNAの配列  
XM\_ 予測されたmRNA  
NR\_ RNAの配列  
XR\_ 予測されたRNAの配列



# NCBI横断検索 - GQuery

NCBI Resources How To Sign in to NCBI

Search NCBI databases

Rous sarcoma virus

About 18,069 search results for "Rous sarcoma virus"

Literature		Genes	
Books	150	EST	17
MeSH	12	Gene	37
NLM Catalog	14	GEO DataSets	8
PubMed	4,250	GEO Profiles	0
PubMed Central	11,923	HomoloGene	0
<b>Health</b>		PopSet	2
ClinVar	127	UniGene	3
dbGaP	5	<b>Proteins</b>	
GTR	0	Conserved Domains	11
MedGen	2	Protein	694
OMIM	4	Protein Clusters	0
PubMed Health	0	Structure	175
<b>Genomes</b>		<b>Chemicals</b>	
Assembly	0	BioSystems	78
BioProject	1	PubChem BioAssay	75
BioSample	0	PubChem Compound	0
Clone	0	PubChem Substance	1
dbVar	3		
Epigenomics	0		
Genome	1	<a href="http://www.ncbi.nlm.nih.gov/">http://www.ncbi.nlm.nih.gov/</a>	
GSS	0		
Nucleotide	463		
Probe	10		
SNP	0		
SRA	0		
Taxonomy	1		

# NCBIゲノムページ

NCBI Resources How To Sign in to NCBI

Genome Genome txid11886[Organism:noexp] Search Save search Limits Advanced Help

Display Settings: Overview Send to: Tools

Organism Overview ; Genome Assembly and Annotation report [1]

ID: 5120

## Rous sarcoma virus

Rous sarcoma virus RefSeq Genome

Lineage: Viruses[3915]; Retro-transcribing viruses[118]; Retroviridae[55]; Orthoretrovirinae[44]; Alpharetrovirus[7]; Rous sarcoma virus[1]

• Rous sarcoma virus

Submitter: NLM, NIH, USA, Bethesda

Status: Complete

BioProjects: PRJNA14978

Type	Name	RefSeq	INSDC	Size (Kb)	GC%	Protein	Gene
	-	NC_001407.1	AF033808.1	9.39	54.1	4	4

Genome Region

Go to nucleotide Graphics FASTA GenBank

Search See more...

リンクからゲノム配列を取得可能

# GOLD: Genome OnLine Database

<http://www.genomesonline.org/>

- ◆ 1995年のHaemophilus influenzaeゲノムの解読以来、現在までにゲノムが解読された／解読中の生物を集めたDB
- ◆ メタゲノムプロジェクト（ある環境にいる微生物のゲノムをまとめて読む）もある
  - 腸内細菌、皮膚常在菌、海水、排水、etc...

**Welcome to the Genomes OnLine Database**

**GOLD Release v.5**

**Studies** 22747

**Biosamples** 54174

**Sequencing Projects** 54174

**Analysis Projects** 35725

**Studies**

- Metagenomic [462](#)
- Non-Metagenomic [22285](#)

**Biosamples**

- [Classification](#)
- Ecosystems
- Host-associated [1608](#)
- Engineered [229](#)
- Environmental [3055](#)

**Projects**

- [Complete Projects 6398](#)
- [Permanent Drafts 17232](#)
- [Incomplete Projects 24584](#)
- [Targeted Projects 972](#)

**Organisms**

- Organisms [49281](#)
- Archaea [884](#)
- Bacteria [35667](#)
- Eukarya [8356](#)

# UCSC Genome Browser

<https://genome.ucsc.edu/>

UCSC Genome Bioinformatics

Genomes Blat - Tables - Gene Sorter - PCR - VisiGene - Session - FAQ - Help

Genome Browser ENCODE Neandertal Blat Table Browser Gene Sorter In Silico PCR Genome Graphs Galaxy VisiGene Utilities Downloads Release Log

Human (*Homo sapiens*) Genome Browser Gateway

The UCSC Genome Browser was created by the [Genome Bioinformatics Group of UC Santa Cruz](#). Software Copyright (c) The Regents of the University of California. All rights reserved.

group genome assembly position search term

Mammal Human Feb. 2009 (GRCh37/hg19) chr1:7,709,963-7,709,977 enter position, gene symbol or search terms submit

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> zoom in 1.5x 3x 10x base ZOOM out 1.5x 3x 10x 100x

chr1:7,709,963-7,709,977 15 bp. enter position, gene symbol or search terms go

Scale chr1: 7,709,963 7,709,964 7,709,965 7,709,966 7,709,967 7,709,968 7,709,969 7,709,970 7,709,971 7,709,972 7,709,973 7,709,974 7,709,975 7,709,976 7,709,977 T

UCSC Genes (RefSeq, GenBank, CDDS, Rfam, tRNAs & Comparative Genomics)

RefSeq Genes

Sequences SNPs

Human mRNAs

Spliced ESTs

Layered H3K27ac

DNase Clusters

Txn Factor ChIP

100 Vert. Cons

Gaps

Human, Rhesus, Mouse, Dog, Elephant, Chicken, X\_tropicalis, Zebrafish, Lamprey

Common SNPs(138)

RepeatMasker

5 bases

Publications: Sequences in Scientific Articles

Human mRNAs from GenBank

Human ESTs That Have Been Spliced

H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE

Digital DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE

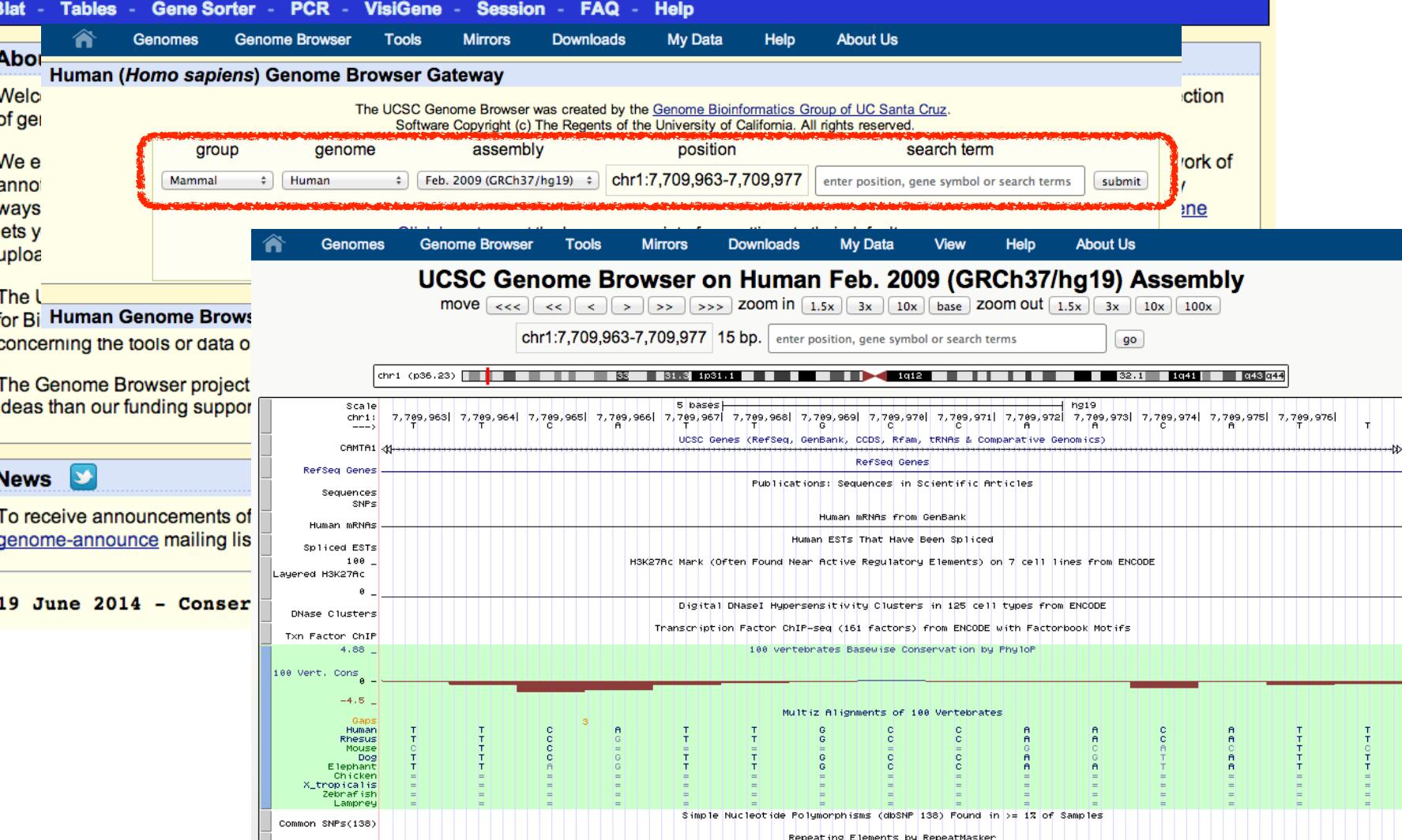
Transcription Factor ChIP-seq (161 factors) from ENCODE with Factorbook Motifs

100 vertebrates Basewise Conservation by PhyloP

Multiz Alignments of 100 Vertebrates

Simple Nucleotide Polymorphisms (dbSNP 138) Found in >= 1% of Samples

Repeating Elements by RepeatMasker



# Ensembl

---



- ◆ Wellcome Trust Sanger InstituteとEuropean Bioinformatics Institute (EBI)が共同開発しているゲノムブラウザ
  - さまざまなアノテーション情報をゲノム上で見る（ブラウズ）することが可能
  - 脊椎動物を中心
  - EnsemblBacteria, EnsemblFungi, EnsemblMetazoa, EnsemblPlants, EnsemblProtistsというのもある

# 代表的なモデル生物のゲノムデータベース

## ◆ ヒト H-InvDB Annotated Human Gene Database

- H-InvDB: H-Invitational Database
- <http://www.h-invitational.jp/>

## ◆ マウス

- MGI: Mouse Genome Informatics
- <http://www.informatics.jax.org/>

## ◆ ショウジョウバエ



GTGGCAATCCTAAGATAGCCAAATTATTATTGTCAGATACTCAC  
AGGAAACATGGAGATTCCTTGGAGTGTTTCTAAATCACTGAAATT  
TCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT  
ATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT  
ATAATAAAAACAAACAAGTGAAAGAACGCCGGGGCATCTTCAATAGA

- FlyBase: FlyBase  
(A Database of Drosophila Genes & Genomes)
- <http://flybase.org/>

## ◆ 線虫

- WormBase
- <http://www.wormbase.org/>

## ◆ ゼブラフィッシュ ZFIN

- ZFIN: The Zebrafish Model Organism Database
- <http://zfin.org/>

## ◆ シロイヌナズナ

- TAIR: The Arabidopsis Information Resource
- <http://www.arabidopsis.org/>

## ◆ 酵母 SGD *Saccharomyces* GENOME DATABASE

- SGD: Saccharomyces Genome Database
- <http://www.yeastgenome.org/>

## ◆ シアノバクテリア

- CyanoBase: Genome database for Cyanobacteria
- <http://genome.microbedb.jp/cyanobase>

# 実習

Ensemblでヒトアセトアルデヒドデヒドロゲナーゼ2  
(ALDH2) 周辺のアノテーション情報を眺める

# 実習1：Ensemblを使ってみる

<http://www.ensembl.org/>

生物種を選択して ALDH2 で検索

Ensembl ASIA

BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Search ✓ All species

Favourite species

Human (circled)  
Mouse  
Zebrafish

Alpaca  
Anole lizard  
Armadillo  
Bushbaby  
C.intestinalis  
C.savignyi  
Caenorhabditis elegans  
Cat  
Cave fish  
Chicken  
Chimpanzee  
Chinese softshell turtle  
Cod  
Coelacanth  
Cow  
Dog  
Dolphin  
Duck  
Elephant  
Ferret  
Flycatcher  
Fruitfly  
Fugu  
Gibbon  
Gorilla  
Gymnophius  
Hedgehog

for ALDH2 (circled)

Go

200000 or coronary heart disease

ENCODE data in Ensembl

Variant Effect Predictor VeIP

Gene expression in different tissues

Find SNPs and other variants for my gene

Retrieve gene sequence

Compare genes across species

Use my own data in Ensembl

Learn about a disease or phenotype

What's New in Release 75 (February 2014)

- New VEP interface
- New 'Age of base' track for human
- New GENCODE basic renderer for human and mouse

Full details of this release

More release news on our blog →

Latest blog posts

- 27 Jun 2014: [ensembldb.ensembl.org maintenance, 1st July 2014](#)
- 13 Jun 2014: [What's coming in Ensembl release 76](#)
- 04 Jun 2014: [Adding value to the human assembly: manual annotation on human GRCh38 released](#)

Did you know...?

"seq" : "MAA  
LLDKFGGEHNPP

Take a REST! Try our [REST API](#) (beta version) for quick access to Ensembl data.

15

(cc) BY ©2014 統合データベース講習会 Licensed Under CC 表示 2.1

# 対象遺伝子の選択

Ensembl ASIA

BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Human (GRCh37) ▾

Current selection:

< all Species

Only searching Human

Restrict category to:

Gene	1
Transcript	4
Variation	980
Somatic Mutation	74
GeneTree	1
Marker	1
ProbeFeature	34
Protein Family	1

Per page:

10 25 50 100

Layout:

Standard Table

Only searching Human ▾ ALDH2

1096 results match ALDH2 when restricted to species: Human ✖

Did you mean... ▾

**ALDH2 (Human Gene)**

**ENSG00000111273** 12:112204691-112247782:1  
Aldehyde dehydrogenase 2 family (mitochondrial) [Source:HGNC Symbol;Acc:404] **ALDH2** (UniProtKB Gene Name), with a synonym of ALDM, is associated with Gene ENSG00000111275.  
[Variation table](#) • [Location](#) • [Regulation](#) • [Orthologues](#) • [Gene tree](#)

**ALDH2-003 (Human Transcript)**  
**ENST00000416293** 12:112204691-112247442:1  
Aldehyde dehydrogenase 2 family (mitochondrial) [Source:HGNC Symbol;Acc:404] **ALDH2-003** (Vega transcript) is associated with Transcript ENST00000416293.  
[Location](#) • [cDNA seq.](#) • [Variation table](#) • [Protein seq.](#) • [Population](#) • [Protein](#)

**ALDH2-001 (Human Transcript)**  
**ENST00000261733** 12:112204726-112247782:1  
Aldehyde dehydrogenase 2 family (mitochondrial) [Source:HGNC Symbol;Acc:404] **ALDH2-001** (Vega transcript) is associated with Transcript ENST00000261733.  
[Location](#) • [cDNA seq.](#) • [Variation table](#) • [Protein seq.](#) • [Population](#) • [Protein](#)

**ALDH2-005 (Human Transcript)**  
**ENST00000548536** 12:112204752-112247434:1  
Aldehyde dehydrogenase 2 family (mitochondrial) [Source:HGNC Symbol;Acc:404] **ALDH2-005** (Vega transcript) is associated with Transcript ENST00000548536.  
[Location](#) • [cDNA seq.](#) • [Variation table](#) • [Protein seq.](#) • [Population](#) • [Protein](#)

# サマリページの表示

ゲノムブラウザで表示

Ensembl ASIA

Human (GRCh37) Location: 12:112,204,691-112,247,782 Gene: ALDH2

Login/Register

Search Human...

Gene-based displays

- Summary
- Splice variants (4)
- Transcript comparison
- Supporting evidence
- Sequence
  - Secondary Structure
  - External references
  - Regulation
  - Expression
- Comparative Genomics
  - Genomic alignments
  - Gene tree (image)
    - Gene tree (text)
    - Gene tree (alignment)
    - Gene gain/loss tree
  - Orthologues (60)
  - Paralogues (11)
  - Protein families (2)
- Phenotype
- Genetic Variation
  - Variation table
  - Variation image
  - Structural variation
- External data
  - Personal annotation
- ID History
  - Gene history

Configure this page

Add your data

Export data

Bookmark this page

Share this page

Gene: ALDH2 ENSG00000111275

Description: aldehyde dehydrogenase 2 family (mitochondrial) [Source:HGNC Symbol;Acc:404]

Location: Chromosome 12: 112,204,691-112,247,782 forward strand.

INSDC coordinates: chromosome:GRCh37:CM000674.1:112204691:112247782:1

Transcripts: This gene has 4 transcripts (splice variants) Show transcript table

Summary

Name: ALDH2 (HGNC Symbol)

CCDS: This gene is a member of the Human CCDS set: [CCDS55885](#), [CCDS9155](#)

RefSeq: Overlapping RefSeq Gene ID [217](#) matches and has similar biotype of protein\_coding

Ensembl version: ENSG00000111275.8

Gene type: Known protein coding

Prediction Method: Annotation for this gene includes both automatic annotation from Ensembl and [Havana](#) manual curation, see [article](#).

Alternative genes: This gene corresponds to the following database identifiers:  
Havana gene: [OTTHUMG00000169603](#) (version 2)

Go to Region in Detail for more tracks and navigation options (e.g. zooming)

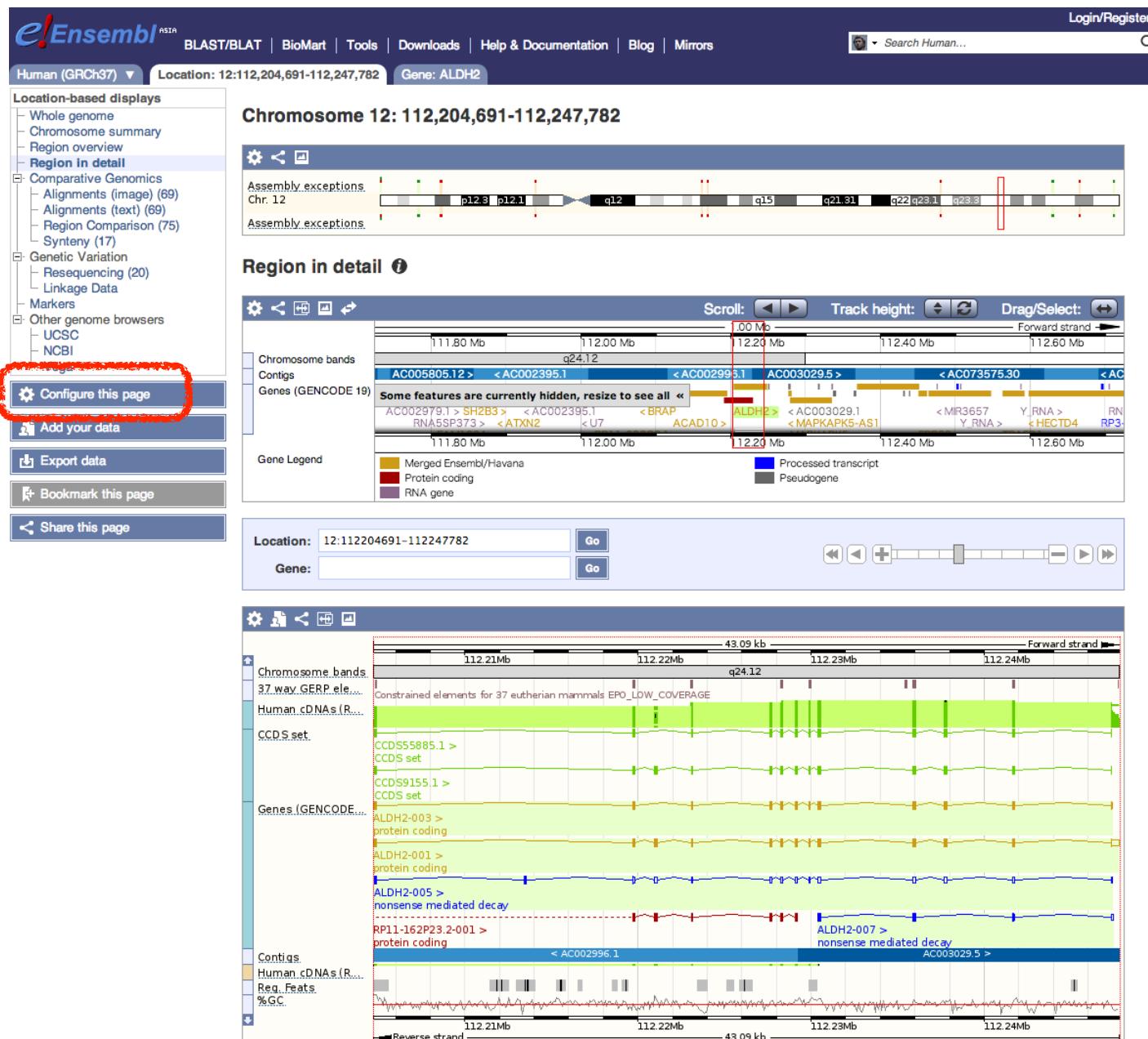
Genes (GENCODE): RP11-162P23.2-001 > protein coding, ACAD10-002 > protein coding, ACAD10-001 > protein coding, ACAD10-004 > processed transcript, ACAD10-017 > retained intron, ALDH2-003 > protein coding, ALDH2-001 > protein coding, ALDH2-005 > nonsense mediated decay, ALDH2-007 > nonsense mediated decay, RP3-462E2.3-001 > lincRNA

Contigs: < AC002996.1 > AC003029.5 >

Gene Legend: protein coding, processed transcript, RNA gene

©2014 統合データベース講習会 Licensed Under CC 表示 2.1

# ゲノムブラウザでの表示



# 表示するアノテーションの指定

新たに表示させる項目を選択

The screenshot shows the Ensembl genome browser interface for Human (GRCh37) at location 12:112,204,691-112,247,782, specifically for the gene ALDH2. The main navigation bar includes links for BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, Blog, and Mirrors. A search bar at the top right allows searching for 'Human...'. The left sidebar provides a hierarchical tree view of genomic regions and other data types. The central panel is titled 'Active tracks' and contains several sections:

- Sequence and assembly:** Contigs, Sequence, Primary assembly mapping, GRC alignment import.
- Genes and transcripts:** GENCODE 19 genes, CCDS set.
- mRNA and protein alignments:** Human cDNAs (RefSeq/ENA).
- Regulation:** Reg. Feats.
- Comparative genomics:** Constrained elements for 37 eutherian mammals EPO\_LOW\_COVERAGE.
- Information and decorations:** %GC, Chromosome bands, Assembly exceptions, Scale bar, Ruler, Variation Legend, Structural Variation Legend, Alignment Difference Legend, Gene Legend, Reg. Features Legend, Methylation Legend.

On the right side of the active tracks panel, there is a vertical list of track icons, each with a star rating and an information icon (i). Some tracks have a checkmark indicating they are currently selected or active. A red box highlights the 'Active tracks' section and the sidebar menu.

# 表示するアノテーションの指定

設定後、✓マークをクリックすると自動更新される

The screenshot shows the Ensembl genome browser interface for the Human (GRCh37) genome. The URL is [Human \(GRCh37\)](#) and the location is [12:112,204,691-112,247,782](#). The gene being viewed is [ALDH2](#).

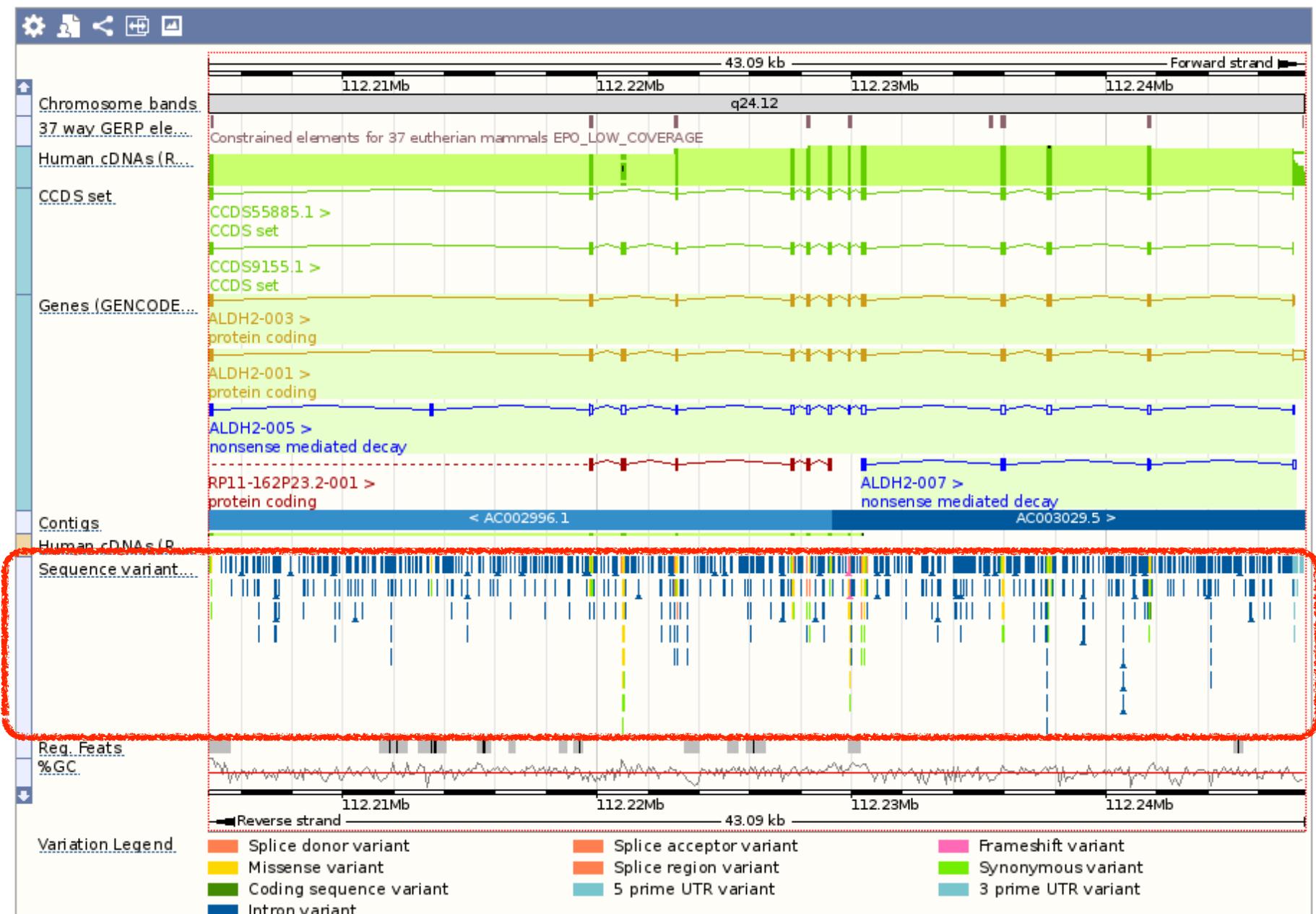
The left sidebar contains a navigation tree with categories like Location, Active tracks, Favourite tracks, Track order, and Variation. The Variation category is highlighted with a red box.

The main content area is titled "Variation" and contains a "Enable/disable all dbSNP" section. A dropdown menu is open, showing options: "Sequence variants (dbSNP and all other sources)" (disabled), "Normal (collapsed for windows over 200kb)" (selected and highlighted with a red box), and "Collapsed". Below this is another section titled "Enable/disable all Phenotype and curated variants" which is currently empty.

On the right side, there is a vertical track list with various genomic features, each with a star icon and an information icon (i). Some tracks are collapsed, indicated by a minus sign.

At the top right, there are links for "Login/Register" and a search bar labeled "Search Human...".

# 選択したアノテーション情報が表示される



# 自分のデータをゲノムブラウザ上に表示

The screenshot shows the Ensembl genome browser interface for Human (GRCh37) at the location 12:112,204,691-112,247,782, specifically for the gene ALDH2. The top navigation bar includes links for BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, Blog, and Mirrors. A search bar is also present.

The main content area displays the Chromosome 12 view with the ALDH2 gene highlighted. Below this, a "Region in detail" panel is shown. On the left, a sidebar provides various options like "Configure this page", "Add your data" (which is highlighted with a red box), "Export data", "Bookmark this page", and "Share this page".

A central dialog box titled "Configure Region Image" is open, showing a "Custom Data" section. Under "Data format", a dropdown menu is open, listing several file types: BAM, BED, bedGraph, BigBed, BigWig, TrackHub, GBrowse, GFF (which is selected and highlighted with a blue box), GTF, PSL, VCF, VEP, and WIG. Below the dropdown, there are fields for "Name for this data (optional)", "Species", "Assembly", "Type", "Paste data", "Or choose file:", and "Or provide file URL:". At the bottom right of the dialog is an "Upload" button.

To the right of the browser interface, a large text block reads: "既存のアノテーションと自分のデータをゲノム上で重ね合わせて表示することが可能" (It is possible to overlap existing annotations and your own data on the genome).

# BioMart

---



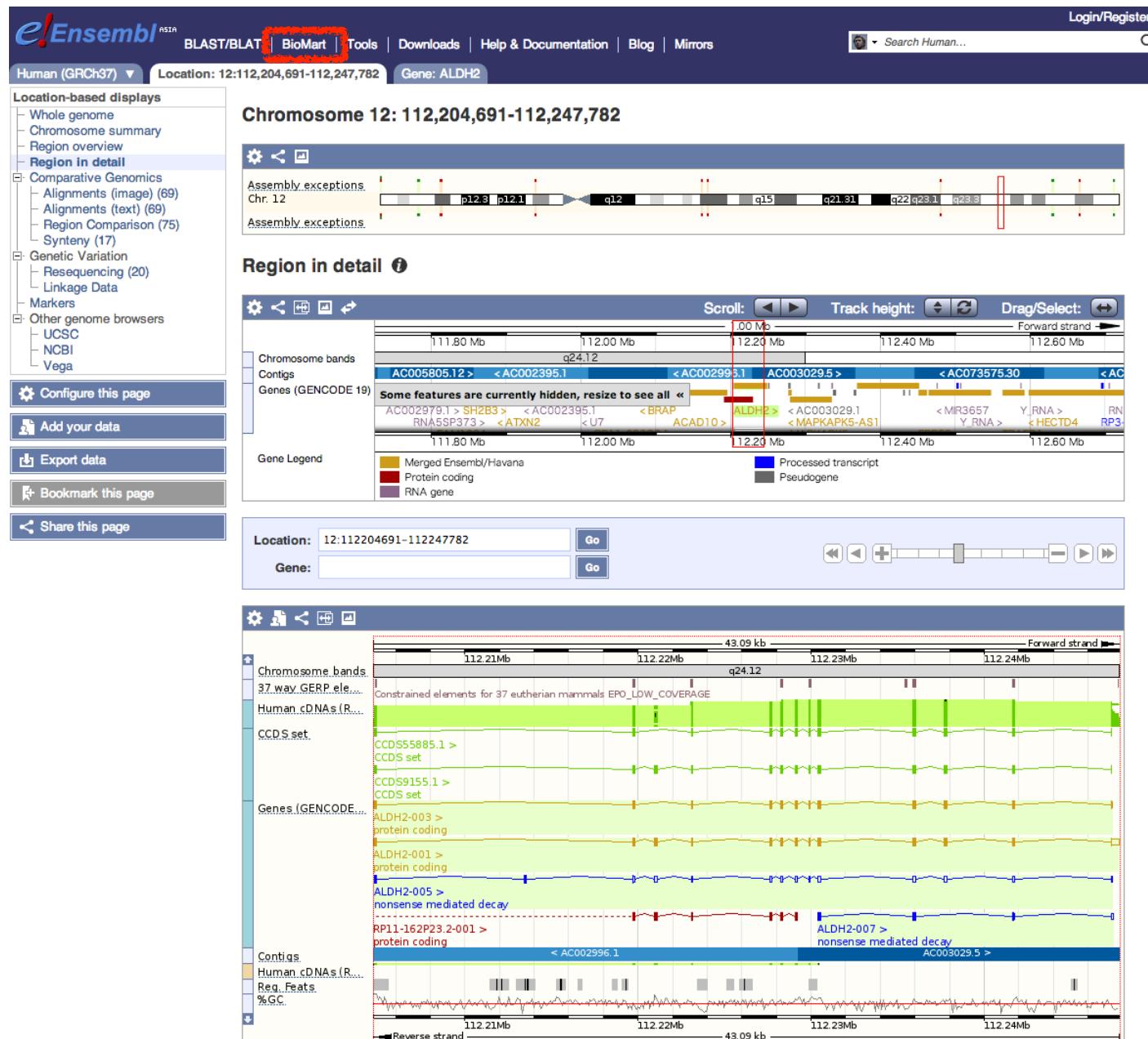
<http://www.biomart.org/>

- ◆ Ontario Institute for Cancer Research (OICR)  
と European Bioinformatics Institute (EBI)が共同開発しているデータ管理システム
- ◆ さまざまなデータベースのデータ取得インターフェイスとして利用されている
  - Ensembl, UniProt, WormBase, KazusaMartなど

# 実習

BioMartを使って、ヒトゲノムの22番染色体にある遺伝子の  
転写開始点から上流の配列1kbをFASTA形式で取得する

# 実習2：BioMartを使ってEnsemblからデータ取得



# データセットを選択

The screenshot shows the Ensembl ASIA web interface. At the top, there is a navigation bar with links for BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, and Blog. On the right side of the navigation bar is a search bar labeled "Search all species..." with a magnifying glass icon. To the right of the search bar is a "Login/Register" link.

Below the navigation bar, there is a toolbar with buttons for "New", "Count", and "Results". To the right of the toolbar are links for "URL", "XML", "Perl", and "Help".

The main content area has two sections. The left section is titled "Dataset" and contains the message "[None selected]". A dropdown menu is open over this section, titled "✓ - CHOOSE DATABASE -". The options listed are Ensembl Genes 75, Ensembl Variation 75, Ensembl Regulation 75, Vega 55, and PRIDE (EBI UK). The option "Ensembl Genes 75" is highlighted with a blue background.

The right section also has a "Dataset" title and "[None selected]" message. A dropdown menu is open over this section, also titled "✓ - CHOOSE DATABASE -". The options listed are Ensembl Genes 75, Danio rerio genes (Zv9), Gallus gallus genes (Galgal4), Homo sapiens genes (GRCh37.p13) (which is highlighted with a blue background), Mus musculus genes (GRCm38.p2), Rattus norvegicus genes (Rnor\_5.0), and a long list of other species' gene databases including Ailuropoda melanoleuca genes (ailMe1), Anas platyrhynchos genes (BGl\_duck\_1.0), Anolis carolinensis genes (AnoCar2.0), Astyanax mexicanus genes (AstMex102), Bos taurus genes (UMD3.1), Caenorhabditis elegans genes (WBcel235), Callithrix jacchus genes (C\_jacchus3.2.1), Canis familiaris genes (CanFam3.1), Cavia porcellus genes (cavPor3), Chloepus hoffmanni genes (choHof1), Ciona intestinalis genes (KH), Ciona savignyi genes (CSAV2.0), Dasypus novemcinctus genes (Dasnov3.0), Dipodomys ordii genes (dipOrd1), Drosophila melanogaster genes (BDGP5), and Echinops telfairi genes (TENREC).

At the bottom left of the page is a Creative Commons Attribution (CC BY) license logo. At the bottom center is the text "©2014 統合データベース".

# 取得したい遺伝子セットを指定

The screenshot shows the Ensembl ASIA BioMart interface. At the top, there is a navigation bar with links for BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, and Blog. On the right side of the top bar are 'Login/Register' and a search bar labeled 'Search all species...' with a magnifying glass icon.

The main interface has three tabs at the top: 'New', 'Count', and 'Results'. Below these are four download links: URL (with a star icon), XML, Perl, and Help. The 'Dataset' section on the left shows 'Homo sapiens genes (GRCh37.p13)' selected. A red box highlights the 'Filters' link under the dataset dropdown. The 'Attributes' section lists Ensembl Gene ID and Ensembl Transcript ID. Another red box highlights the 'Filters' link under the attributes dropdown.

The central area contains a search form with dropdown menus for 'Ensembl Genes 75' and 'Homo sapiens genes (GRCh37.p13)'. Below this is another set of tabs: 'New', 'Count', 'Results', 'URL', 'XML', 'Perl', and 'Help'. To the right of this is a section titled 'Please restrict your query using criteria below' with various filter options, each preceded by a checkbox:

- REGION:
- GENE:
- PHENOTYPE:
- TRANSCRIPT EVENT:
- GENE ONTOLOGY:
- EXPRESSION:
- MULTI SPECIES COMPARISONS:
- PROTEIN DOMAINS AND FAMILIES:
- VARIATION:

フィルタ機能で遺伝子を絞り込む

27



©2014 統合データベース講習会 Licensed Under CC 表示 2.1

# 取得したい遺伝子セットを指定

22番染色体を指定（複数の条件を組み合わせることも可能）  
Countで絞りこまれた遺伝子数を表示できる

The screenshot shows the Ensembl ASIA web interface. At the top, there is a navigation bar with links for BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, and Blog. On the right side of the bar are 'Login/Register' and a search bar labeled 'Search all species...' with a magnifying glass icon. Below the navigation bar, there is a toolbar with buttons for 'New', 'Count' (which is highlighted with a red box), and 'Results'. To the right of the toolbar are links for 'URL', 'XML', 'Perl', and 'Help'. The main content area has a heading 'Please restrict your query using criteria below'. On the left, there is a sidebar with sections for 'Dataset 1263 / 64162', 'Genes' (highlighted with a red box), 'Homo sapiens genes (GRCh37.p13)', 'Filters' (highlighted with a red box), 'Chromosome: 22' (highlighted with a red box), 'Attributes' (highlighted with a red box), 'Ensembl Gene ID', and 'Ensembl Transcript ID'. In the main content area, under 'REGION:', there is a checked checkbox for 'Chromosome' and a dropdown menu listing chromosomes from 5 to Y, with '22' selected. There are also checkboxes for 'Base pair', 'Gene Start (bp)' with a value of 1, and 'Gene End (bp)' with a value of 100000000. The 'Attributes' section is at the bottom of the sidebar.

Attributesで取得する情報を指定する

# 取得したいデータの種類を指定

Ensembl ASIA

BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog

Login/Register

Search all species...

New Count Results URL XML Perl Help

Please select columns to be included in the output and hit 'Results' when ready

Features  Homologs  
 Structures  Variation  
 Transcript Event  Sequences

GENE:  
EXTERNAL:  
EXPRESSION:  
PROTEIN DOMAINS AND FAMILIES:

New Count Results URL XML Perl Help

Please select columns to be included in the output and hit 'Results' when ready

Features  Homologs  
 Structures  Variation  
 Transcript Event  Sequences

SEQUENCES:  
Header Information

(cc) BY ©2014 統合データベース

# 取得したいデータの種類を指定

遺伝子の上流域1000bpを指定

Ensembl ASIA

BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog

Login/Register

New Count Results URL XML Perl Help

Dataset 1263 / 64162  
Genes  
Homo sapiens genes (GRCh37.p13)

Filters  
Chromosome: 22

Attributes  
Ensembl Gene ID  
Ensembl Transcript ID  
Flank (Gene)  
Upstream flank [1000]

SEQUENCES:  
Sequences (max 1)

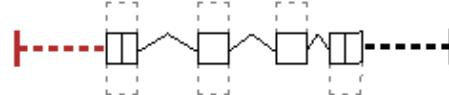
Unspliced (Transcript)  
Unspliced (Gene)  
Flank (Transcript)  
**Flank (Gene)**  
Flank-coding region (Transcript)  
Flank-coding region (Gene)

5' UTR  
3' UTR  
Exon sequences  
cDNA sequences  
Coding sequence  
Protein

Upstream flank  
 Upstream flank

Downstream flank  
 Downstream flank

Header Information



Search all species... 

# 取得したいデータの種類を指定

ヘッダ情報（FASTAのコメント行）の指定

The screenshot shows the Ensembl BioMart interface for specifying data types. The 'Results' tab is selected. On the left, a sidebar shows a dataset of 1263 genes from Homo sapiens (GRCh37.p13) on Chromosome 22. Under 'Attributes', 'Associated Gene Name' is selected. In the main panel, under 'Gene Information', 'Associated Gene Name' is checked and highlighted with a red box. Under 'Transcript Information', 'Ensembl Transcript ID' is also highlighted with a red box. Other options like 'Ensembl Gene ID' and 'Ensembl Protein ID' are available but not checked.

Dataset 1263 / 64162  
Genes  
Homo sapiens genes (GRCh37.p13)  
Filters  
Chromosome: 22  
Attributes  
Ensembl Gene ID  
Flank (Gene)  
Upstream flank [1000]  
Associated Gene Name  
Dataset  
[None Selected]

Header Information

Gene Information

Ensembl Gene ID  
 Description  
 Associated Gene Name  
 Associated Gene DB  
 Chromosome Name

Gene Start (bp)  
 Gene End (bp)  
 Gene Biotype  
 Ensembl Protein Family ID(s)

Transcript Information

CDS start (within cDNA)  
 CDS end (within cDNA)  
 5' UTR Start  
 5' UTR End  
 3' UTR Start  
 3' UTR End

Ensembl Transcript ID  
 Ensembl Protein ID  
 Transcript Biotype  
 Strand  
 Transcript Start (bp)  
 Transcript End (bp)

Exon Information

CDS Length  
 CDS Start

Exon Rank in Transcript  
 phase

Resultsで取得データを表示

# データのダウンロード

Ensembl ASIA

BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog

Login/Register

New Count Results URL XML Perl Help

Dataset 1263 / 64162

Genes

Homo sapiens genes  
(GRCh37.p13)

Filters

Chromosome: 22

Attributes

Ensembl Gene ID

Flank (Gene)

Upstream flank [1000]

Associated Gene Name

Dataset

[None Selected]

Export all results to

File

Compressed file (.gz)

Compressed web file (notify by email)

FASTA

Unique results only

Go

View

10 rows as FASTA Unique results only

```
>ENSG00000008735 | MAPK8IP2
AGGCAGCCGCCACCACGCCGGCTAATTTTTGTATTTAGTAAAGACGGGGTTTCAC
CATGTTGGCCAGGATGGTCTCTGGTCTCGATCTCTTGACGTCGTGATCCACCCGCCCTGG
CCTCTCAAAGTGGTGGGATTGCAGGCAGCCACCGCAGCCACCCGGTCTCTATTT
TAGAAAAGAAAAAAAGTGTCTGCAGTCTCTCCCTAGTTCTGCGCCCTCAACCCCTGATCTC
CCGCCTCCTCCCACAGCCACCGTCTCGAGCTCAGCTTAACCATGGCCCTCCCCACAGGG
CTCCATCCCAAGCCAGAGCGGGTCAGTCCCTACACAGGGCCAGCAACACGGTGCCCAA
GAAAAAAGTTTCCAGTGGGACGCATGTCGCCCTCCGTGGGTGCTAACTCTCCAGTGCC
CGTCGGTACGTCCGTCCGGCTGTGAGTCTACTGAGTGCTCAAAGTCCACTTTAGCC
CCAGCCAGCATCCTCTGAGTACCTTGTGCTCGATTCTGAATCCAGCCCTGCTGGCGACC
CTCGTGCAGACGTCCCCGTGCAGCCCCGGTCAGCACCCGTGGGACAGCTCCGATCAGC
ACCCGACGGCGGACAGCTCTCCCGCCTCCCGCCTCGCCCCGCCACCCGCTCCGGGCC
GCACCCGGGTTAGGGTTCTGGGGGATCCTGGGCGGGCGGGCAGCTCCGGGCCG
CCCGGCCTCTCGCCCCAGAGGAAACAGCCCACGGCGAACCCGAGGGCCGGACCCGGCG
CTGCCTCCTCGGCCTCCCGCCCCGGCGGGTCCTCCAGGGCGGGCCGACGCCACCCG
CCCGCCCTCTCCCTCCCGCCCCCTCCCCGGGAAAAATCCTCATCGCACGCCGGCCCTCC
CCAACCAGCGCCCCGGCAGTGAAGCCGCCCTCTCCCAACCAGCGTCCCCGGCCGCGT
CCCTCCCCCGCCTCCCCGGCCCTCCCCGGCCGCG
>ENSG00000015475 | BID
CCATCTGACGGTTCCCTGCTCACAGGATTTGTTAGCCCCCAGAGTGCCAG
```

# ゲノム配列・転写産物配列の高速検索

---

## ◆ GGRNA (ググるな)

- <http://ggrna.dbcls.jp/ja/>
- RefSeqのトランスクリプト配列を高速検索

## ◆ GGGenome (ゲゲゲノム)

- <http://gggenome.dbcls.jp/ja/>
- ゲノム配列を高速検索

# 実習

GGRNA、GGGenomeを使ってみる

<http://ggrna.dbcls.jp/ja/>

<http://togotv.dbcls.jp/20120124.html>

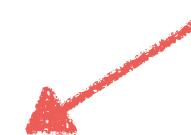
## 統合遺伝子検索

# GGRNA

ver.2

[Help](#) | [Advanced search](#) | [English](#) | [旧バージョン](#)

生物種を選択



検索

Zoo (All organisms in RefSeq)

遺伝子をGoogleのように検索できるサイトです。 [NCBI RefSeq](#) の transcript を全文検索します。

検索例：

- 「[homeobox](#)」「[claudin](#)」..... フリーワード検索
- 「["RNA interference"](#)」..... ダブルクオートで囲ってフレーズ検索
- 「[Argonaute "PAZ domain"](#)」..... Argonaute かつ "PAZ domain" のAND検索
- 「[NM\\_001518](#)」「[10579](#)」..... RefSeq IDやGene IDなど各種IDから検索
- 「[symbol:VIM](#)」..... 遺伝子名(symbolまたはsynonym)から検索
- 「[ref:Naito](#)」..... 文献情報のなかからフリーワード検索
- 「[1552311\\_a\\_at](#)」..... マイクロアレイのプローブIDから塩基配列を検索
- 「[aa:KDEL](#)」..... アミノ酸配列を検索
- 「[caagaagagattq](#)」..... 塩基配列を検索
- 「[comp:caagaagagattq](#)」..... 相補鎖を検索
- 「[iub:aqgtcannrtgacct](#)」..... N, R, Y 等のあいまいな塩基を含む塩基配列を検索
- 「[詳細な使い方](#)」

キーワード・配列を入力

← 検索例（クリック）

新着情報：

- 2014-05-30 データベースをRefSeq rel. 65 (May, 2014)に更新。
- 2013-07-24 ソースを公開 - [GitHub](#)
- 2013-07-08 GGRNA ver.2公開。全生物種のRefSeqを検索できます。
- 2012-05-29 下記論文の日本語による解説を「DBCLSからの成果発信」に掲載。
- 2012-05-29 GGRNAの論文が[Nucleic Acids Research](#)に掲載されました。
- 過去の新着情報

# GGRNA検索例～マイクロアレイプローブ

プローブに相当する部分が  
緑色で表示される  
(重なると色が濃くなる)

クリックするとRefSeqのエントリが表示される

### **Results:**

トップ50件を表示。検索語に色がつきます。重なると色が濃く表示されます。

position 1592 1634 1650 1698 1717 1783 1807 1812 1955 1972 1973

Synonym: ARMD6; CORD11; ORX; RAXL1

NM\_032753.3 - Homo sapiens (human) - NCBI - UCSC - RefEx(expression)

PREDICTED: Homo sapiens retina and anterior neural fold homeobox 2 (RAX2), transcript variant X1, mRNA. (2414 bp)

**position** 1851 1893 1909 1957 1976 2042 2066 2071 2214 2231 2234

Synonym: ARMD6; CORD11; QRX; RAXL1

XM\_005259662.1 - Homo sapiens (human) - NCBI - UCSC - RefEx(expression)

## Data Export:

下記より最大10000件まで検索結果を取得できます。

- ・タブ区切りテキスト → [表示](#) | [ダウンロード](#)  
エクセル等の表計算ソフトに直接コピペできます
  - ・JSON形式 → [リンク](#) | [ダウンロード](#)

# GGRNA経由で表示したRefseqエントリ

```
//  
variation      /variation_id=1000002  
               /complement(1947)  
               /gene="RAX2"  
               /gene_synonym="ARMD6; CORD11; QRX; RAXL1"  
               /replace="a"  
               /replace="g"  
               /db_xref="dbSNP:138882802"  
variation      /variation_id=1000003  
               /complement(1972)  
               /gene="RAX2"  
               /gene_synonym="ARMD6; CORD11; QRX; RAXL1"  
               /replace="a"  
               /replace="t"  
               /db_xref="dbSNP:8103165"  
variation      /variation_id=1000004  
               /complement(1989)  
               /gene="RAX2"  
               /gene_synonym="ARMD6; CORD11; QRX; RAXL1"  
               /replace="c"  
               /replace="g"  
               /db_xref="dbSNP:6510769"  
polyA_signal   2131..2136  
               /gene="RAX2"  
               /gene_synonym="ARMD6; CORD11; QRX; RAXL1"  
polyA_site     2155  
               /gene="RAX2"  
               /gene_synonym="ARMD6; CORD11; QRX; RAXL1"  
  
ORIGIN  
gcccccaaggcccgctggggcaggtgtcccggtggaaatcgacggaggggctgccgtggcggtggagccatgttccctgagcccgggcgagggggccggcaaccgagggtgggggtctgg  
ggccggggcaggaggagggccccaagaagaaggcaccggaggaaccgcaccacattcaccacacttaccagctgacccagctggagcgggggttcgaggcccttcactacccggatgtgtaca  
gccgtgaggagactggcagccaagggtgcacccatctgagggtgcgcgtgcagggtgtggttccagaaccgcggggcccaagtggcgcggccaggagcgggtggagtcaggctcggtgcgg  
tggcagctcccgagactcccccgaggccccagcgctgcgttgcggccggggccatgtcgctgcgtgcggccctgggtggggccggaccggccgggtggccaggcccttc  
gcctcctggggccggggccgggtgtcaaggcttccttcggccctcatgccttgcgtccacccatccgcagatggcttcgcggccctggaggaggcgtccctggggctgtggccaa  
atgcacagggtctggacaggccctggccgcacggcttgcggccatgcgttgcggccctccggcccttcctggcccaacccgagaaccggggacgtggccctggtagacagccaccacgccttgc  
gcctaggccagggtcatggagcaaccgtggcaggccaccacactggggagcgggaccagagacaggctgtgggttcctgcggccatccgtctccacccatccatgc  
caccctgtctgtggcagcggactggcccccactgtcaggcaggaggtgacccaaatgtcgccatgcgttgcggggacccatccgtctccacccatccatgc  
attggggctggggcctcccgagggttcgagggctgcaggtgtgggtggggaccggctgactctgtaaatgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc  
cccacttcaggctccctggaaataacaaatagcagcagactcccgccggccatgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc  
acctcgggcacagctgtcagccattctatagagagggaaaccggggctttaggcaggaaagcagggtccccaatgcacggccaggagtggtggagctgcgttgcgttgcgttgc  
ccggtectaccgtccgggcacacgcacagggtctggagagagggtgggtccggccagggtgtgggtggccatccagggtggaggcgggctgtatgtgggtgtcgatcg  
accgttctcactgtcccccacagaccccaggccctgtcatgttcccaatgtggaggcatggccagcatctgtctgttcaaccacccagccatgcggccaaagagctctgagaagg  
gtcgccggccgagaaccgcgtggactggcaagcacgcgtggcccaatgtggaggcatggccatgtggaggcatggccatggccatggccatggccatggccatggccatggccatgg  
tgcggcaccaggcaggagctgcagtagctaccctccctgtccacgcctgggtcccaatgcacccaggatcactcccaatgcacccatctctaggcggcacctc  
ctgtggtccaacccccccatcaccatgtcaccgcacaccaggcaggatggggacacggcaggtaagcacaagaaagattttttttaaagactaaaccaggccagtgccgtgg  
tcatgcctgtatccctgtcatcccacgttcccaagaaagctgaggcaagaggatcacttgcgttgcaggcacttgcgttgcaggctgtgtatcactgtcactcc  
tgggtgcacacactgtcatcccacgttcccaagaaagctgaggcaagaggatcacttgcgttgcaggcacttgcgttgcaggctgtgtatcactgtcactcc  
agagccagaccctgtactcaataaaaaaaaatctaaaacaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa  
//
```

# GGRNA：検索タグ一覧

検索語	説明	エイリアス
<u>refid:NM_001518</u>	RefSeq IDから検索。 "." 以降のバージョン番号は無視される。 つまり古い配列 NM_003380.2 を検索しても最新版 NM_003380.3 がhitする。 NM_, XM_, NR_, XR_ で始まる検索語は自動的にrefid:で解釈される。	refseqid: refseq: id:NM_ id:XM_ id:NR_ id:XR_
<u>geneid:10579</u>	Gene IDから検索。 数字のみの検索語は自動的にgeneid:で解釈される。	gene:数字 id:数字
<u>symbol:VIM</u>	symbolから検索。 symbolに部分一致するものを返す。 synonymに部分一致するものも返す。	name: gene:アルファベットを含む単語
<u>aa:KDEL</u>	アミノ酸配列から検索。	
<u>ref:Naito</u> <u>ref:1327585</u>	文献情報のなかからフリーワード検索。 PubMed IDも指定可能。	reference:
<u>probe:1552311_a_at</u> <u>probe:A_23_P101434</u>	アレイのプローブIDから塩基配列を検索。 _at, _stで終わるもの（Affymetrix ID）、A_で始まるもの（Agilent ID）は自動的にprobe:で解釈される。プローブIDが登録されていない場合はフリーワード検索。	probeid:
<u>seq:caagaagagattq</u>	塩基配列から検索。 A,T,G,C,Uのみからなる文字列は自動的にseq:で解釈される。UはTとまったく同じ。 seq1:, seq2:, seq3: とすると、それぞれ1, 2, 3ミスマッチまで許容して検索する。	sequence:
<u>comp:caagaagagattg</u>	相補鎖を検索する。 comp1:, comp2:, comp3: とすると、それぞれ1, 2, 3ミスマッチまで許容して検索する。	complementary:
<u>both:caagaagagattq</u>	入力配列とその相補鎖を検索する。 both1:, both2:, both3: とすると、それぞれ1, 2, 3ミスマッチまで許容して検索する。	bothseq:
<u>iub:aqqtcanrtgacct</u> <u>iubcomp:aqqtcanrtgact</u> <u>iubboth:aqqtcanrtgacct</u>	N, R, Yなどのあいまいな塩基を含む塩基配列を検索。あいまい塩基の表記はIUBコードによる。iubcomp:は相補鎖、iubboth:は入力配列とその相補鎖を検索。	iubseq: → iub:

検索タグを先頭につけることで  
検索フィールドを絞って検索できる  
(RefSeq IDなど自動的に認識可能な  
ものは省略可能)

# GGGenome

<http://gggenome.dbcls.jp/ja/>

<http://togotv.dbcls.jp/20131025.html>

超絶高速ゲノム配列検索

Help | English

GGGenome

許容するミスマッチ/ギャップの数 :  (検索する塩基配列の長さの20%まで)

検索

Human genome, GRCh37/hg19 (Feb, 2009)

▼

検索例 :

- [\[ TTCATTGACAACATT \] ..... 塩基配列を検索](#)
- [詳細な使い方](#)

検索結果へのリンク :

検索例 (クリック)

- [http://GGGenome.dbcls.jp/db/k/sequence\[.format\]\[.download\]](http://GGGenome.dbcls.jp/db/k/sequence[.format][.download])
  - db → hg19, mm10, rn5, susScr3, galGal4, xenTro3, danRer7, ci2, dm3, ce10, TAIR10, rice, sorBic, bmor1, sacCer3, refseq, hs\_refseq, prok, ddbj。省略時は hg19
  - k → 許容するミスマッチ/ギャップの数。あまり大きいとしぼうする。省略時は 0
  - sequence → 塩基配列。大文字・小文字は区別しない
  - format → html, txt, json。省略時は html
  - download → URLの最後に付加すると検索結果をファイルとしてダウンロードできる
- 例1 : <http://GGGenome.dbcls.jp/TTCATTGACAACATT>
  - ヒトゲノム hg19 (省略可) で
  - ミスマッチ/ギャップを許容せず (省略可)
  - **TTCATTGACAACATT** を検索し
  - **html** 形式 (省略可) で結果を返す
- 例2 : <http://GGGenome.dbcls.jp/mm10/2/TTCATTGACAACATTGCGT.txt>
  - マウスゲノム mm10 で
  - 2 ミスマッチ/ギャップまで許容して
  - **TTCATTGACAACATTGCGT** を検索し
  - **txt** 形式 (タブ区切りテキスト) で結果を返す

検索

Human genome, GRCh37/hg19 (Feb, 2009)

▼

配列を入力

生物種を選択

# GGGenome検索結果

超絶高速ゲノム配列検索

[Help](#) | [English](#)

GGGenome

TTCATTGACAACATT

検索

Human genome, GRCh37/hg19 (Feb, 2009)

許容するミスマッチ/ギャップの数 :  (検索する塩基配列の長さの20%まで)

2014-06-30 14:36:28, GGGenome : Human genome, GRCh37/hg19 (Feb, 2009)

## Summary:

ミスマッチを許容して検索

- TTCATTGACAACATT (15)
- AATGTTGTCATGAA (8)
- TOTAL (23)

入力配列

## Results:

相補配列

+鎖および-鎖それぞれ50件まで表示。検索語に色がつきます（ミスマッチ・挿入欠失）。

[chr1:83462476-83462490](#)

▼83462476

CTAATTAAATGAATAAGTATAAGTATATTTCTCTTCATTGACAACATTTTCTCTACTTATTGTAAAGAACATAAT

[chr2:161223115-161223129](#)

▼161223115

ATTGCAGCAAAGAACATTTATATTCTATTTCATTGACAACATTAGCATATAATGCAAATTACTTGGAACTT

[chr3:15289790-15289804](#)

▼15289790

ATTGTAATAGCCAGAAATTGGAAACATATAATTTCATTGACAACATTAAAGATTATAATAGTCATATAATAGTCC

[chr3:84619845-84619859](#)

▼84619845

TTGGTTTGATATATTAAATAGTTGAAACATTTCATTGACAACATTAAAGTTGTTCTAGGCTCTATTGGATAA

[chr4:16134651-16134665](#)

▼16134651

AACGAAAGCGTTGAGGAAGCCTGATAATTTCATTGACAACATTACCATCAACCACACCTTACCATCCTAGGA

[chr4:63031140-63031154](#)

▼63031140

TCTTCAAATATTTCACAGAGTTCACACTCTTATTTCATTGACAACATTATGACATTATGCTGATTGGATCCAGTGAGT



# GGGenomeミスマッチ検索結果

超絶高速ゲノム配列検索

[Help](#) | [English](#)

**GGGenome**

TTCATTGACAAACATT

検索

Human genome, GRCh37/hg19 (Feb, 2009)



許容するミスマッチ/ギャップの数 :  (検索する塩基配列の長さの20%まで)

2014-06-30 14:39:33, GGGenome : Human genome, GRCh37/hg19 (Feb, 2009)

## Summary:

- [TTCATTGACAAACATT \(1007\)](#)
- [AATGTTGTCAATGAA \(943\)](#)
- **TOTAL (1950)**

## Results:

UCSCゲノムブラウザへのリンク

+鎖および-鎖それぞれ50件まで表示 検索語に色がつきます（ミスマッチ・挿入欠失）。

[chr1:7709963-7709977](#)

▼ 7709963

TGCCAGGGAGTGGTTCTGTCAAAGTTAGAA **TTCATTGC** CAACATT CAAATCTCCAGATTTCAGCCTCTTTAA

[chr1:14482580-14482593](#)

▼ 14482580

AAGAAATTTTTAATTTCTTCTTAATTTCTTCATTGAC-ACATT GGTCAATTAGGACCATAATTATTTAATTCC

[chr1:19318674-19318687](#)

▼ 19318674

GGTCACTTCTCTGTCTAGCTCCTGCATAAC **TTC-TTGACAACATT** CAACATAATTGCCATTCCCACCCCTTTGA

[chr1:24735451-24735464](#)

▼ 24735451

TCTATACTAAAAATTGATATTTAATAGAA **TT-ATTGACAACATT** CTTCAAGAAATTAGAGATTCAAATAACA

[chr1:27076473-27076488](#)

▼ 27076473

TGTCTAAAGAGGATGCTTTGAGTAGATAA **TTCATTGAG** CAACATT TATCAAGCATCTGCGTTGTGTCTGG

# GGGenome応用

## 超絶高速ゲノム配列検索 GGGenome

許容するミスマッチ/ギャップの数 :  (検索する塩基配列の長さの20%まで)

検索

Human genome, GRCh37/hg19 (Feb, 2009)

検索例 :

- [\[ TTCATTGACAACATT \]](#) ..... 塩基配列を検索
- [詳細な使い方](#)

検索結果へのリンク :

- [http://GGGenome.dbcls.jp/db/k/sequence\[.format\]\[.download\]](http://GGGenome.dbcls.jp/db/k/sequence[.format][.download])
  - db → hg19, mm10, rn5, susScr3, galGal4, xenTro3, danRer7, ci2, dm3, ce10, TAIR10, rice, sorBic, bmor1, sacCer3, refseq, hs\_refseq, prok, ddbj。省略時は hg19
  - k → 許容するミスマッチ/ギャップの数。あまり大きいとしぼうする。省略時は 0
  - sequence → 塩基配列。大文字・小文字は区別しない
  - format → html, txt, json。省略時は html
  - download → URLの最後に付加すると検索結果をファイルとしてダウンロードできる
- 例1 : <http://GGGenome.dbcls.jp/TTCATTGACAACATT>
  - ヒトゲノム hg19 (省略可) で
  - ミスマッチ/ギャップを許容せず (省略可)
  - TTCATTGACAACATT を検索し
  - html 形式 (省略可) で結果を返す
- 例2 : <http://GGGenome.dbcls.jp/mm10/2/TTCATTGACAACATTGCGT.txt>
  - マウスゲノム mm10 で
  - 2 ミスマッチ/ギャップまで許容して
  - TTCATTGACAACATTGCGT を検索し
  - txt 形式 (タブ区切りテキスト) で結果を返す

URLから直接検索できる

http://GGGenome.dbcls.jp/

+

データベース名

+

許容するミスマッチ数

+

検索配列

+

(出力フォーマット)

+

(ダウンロード)

# GGGenome応用～プライマーセットからPCR增幅断片の長さを求める

超絶高速ゲノム配列検索

[Home](#) | [English](#)

## GGGenome

  Human genome, GRCh37/hg19 (Feb, 2009) 

許容するミスマッチ/ギャップの数 :  (検索する塩基配列の長さの20%まで)

検索例 :

- [ TTCATTGACAACATT ] ..... 塩基配列を検索

ひとこと :

- 塩基配列を高速に検索するサイトです。
- 大文字・小文字は区別しません。
- U は T とまったく同じです。
- ATGCU 以外の塩基 (R,Y等) は N とみなします。
- それ以外の文字は除去して検索されます。
- +鎖および-鎖それぞれ50件まで結果を表示します。
- TXT, JSON形式ではそれぞれ10000件まで取得できます。
- Googleスプレッドシート上で配列検索 ができます。

動画による解説（統合TV）：



# GGGenome応用～プライマーセットからPCR增幅断片の長さを求める

GGGenome primer search ☆

ファイル 編集 表示挿入 表示形式 データ ツール ヘルプ 変更内容をすべてドライブに保存しました

	A	B	C	D	E	F	G	H	I	J	M
2	YN001-F	caatcacccctaccctttatgc	<a href="http://GGGenome.dbcls.jp/rice/caatcacccctaccctttatgc.txt">http://GGGenome.dbcls.jp/rice/caatcacccctaccctttatgc.txt</a>	chr01 + 27707 27731 CCCATCCAGGTACATCGTAG 27607 27831	chr01	+		27707	27731		385
3	YN001-R	ccgctgtgaacaacaatcatgc	<a href="http://GGGenome.dbcls.jp/rice/ccgctgtgaacaacaatcatgc.txt">http://GGGenome.dbcls.jp/rice/ccgctgtgaacaacaatcatgc.txt</a>	chr01 - 28071 28092 ATATTGTTGGTACGAGTA 27971 28192	chr01	-		28071	28092		
4	YN002-F	tgatccaaatagttttat	<a href="http://GGGenome.dbcls.jp/rice/tgatccaaatagttttat.txt">http://GGGenome.dbcls.jp/rice/tgatccaaatagttttat.txt</a>	chr03 + 9682634 9682653 TTTCAAAGATGATAAACCT 9682534 9682753	chr03	+	9682634	9682653			129
5	YN002-R	catgcaaggatttagtgatg	<a href="http://GGGenome.dbcls.jp/rice/catgcaaggatttagtgatg.txt">http://GGGenome.dbcls.jp/rice/catgcaaggatttagtgatg.txt</a>	chr03 - 9682744 9682763 AGTTGTTTATTTTGACTC 9682644 9682863	chr03	-	9682744	9682763			
6	YN003-F	acgtactgtggAACAGTgag	<a href="http://GGGenome.dbcls.jp/rice/acgtactgtggAACAGTgag.txt">http://GGGenome.dbcls.jp/rice/acgtactgtggAACAGTgag.txt</a>	chr06 + 7494240 7494259 TAATTGTTGTAGGCATGAG 7494140 7494359	chr06	+	7494240	7494259			246
7	YN003-R	acccaacctaatactatgaa	<a href="http://GGGenome.dbcls.jp/rice/acccaacctaatactatgaa.txt">http://GGGenome.dbcls.jp/rice/acccaacctaatactatgaa.txt</a>	chr06 - 7494467 7494486 GAAACATAGGACCGGATGC 7494367 7494586	chr06	-	7494467	7494486			
8	YN004-F	atcagattccggcgccgccc	<a href="http://GGGenome.dbcls.jp/rice/atcagattccggcgccgccc.txt">http://GGGenome.dbcls.jp/rice/atcagattccggcgccgccc.txt</a>	chr01 + 27691273 27691292 CACCGGAACGAGACGACAA 27691173 27691392	chr01	+	27691273	27691292			580
9	YN004-R	ggagagatctggggggag	<a href="http://GGGenome.dbcls.jp/rice/ggagagatctggggggag.txt">http://GGGenome.dbcls.jp/rice/ggagagatctggggggag.txt</a>	chr01 - 27691834 27691853 GCCTCCCCAGGCGTCGA 27691734 27691953	chr01	-	27691834	27691853			
10											
11											
12											

入力

Concatenate関数で文字列結合

結果をセルに分割

ImportData関数で結果取得

塩基数計算

# 本日紹介する内容（次世代シーケンスDB）

---

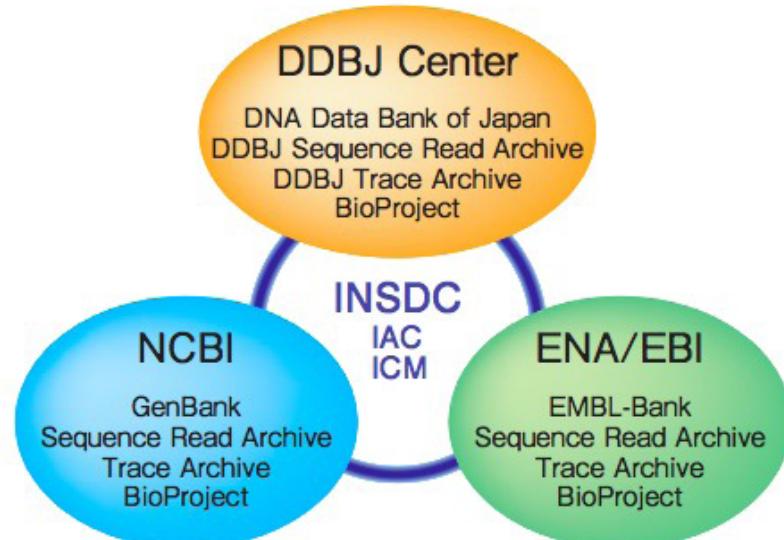
- ◆ 次世代シーケンスデータベースの概要
- ◆ 【実習】 DRAsearch
  - ◆ 次世代シーケンスデータベース検索
- ◆ 【実習】 DDBJ解析パイプライン
  - ◆ ウェブ上で次世代シーケンスデータを解析する
- ◆ Galaxy/MiGAP
  - ◆ 配列データのアノテーション

# Sequence Read Archive

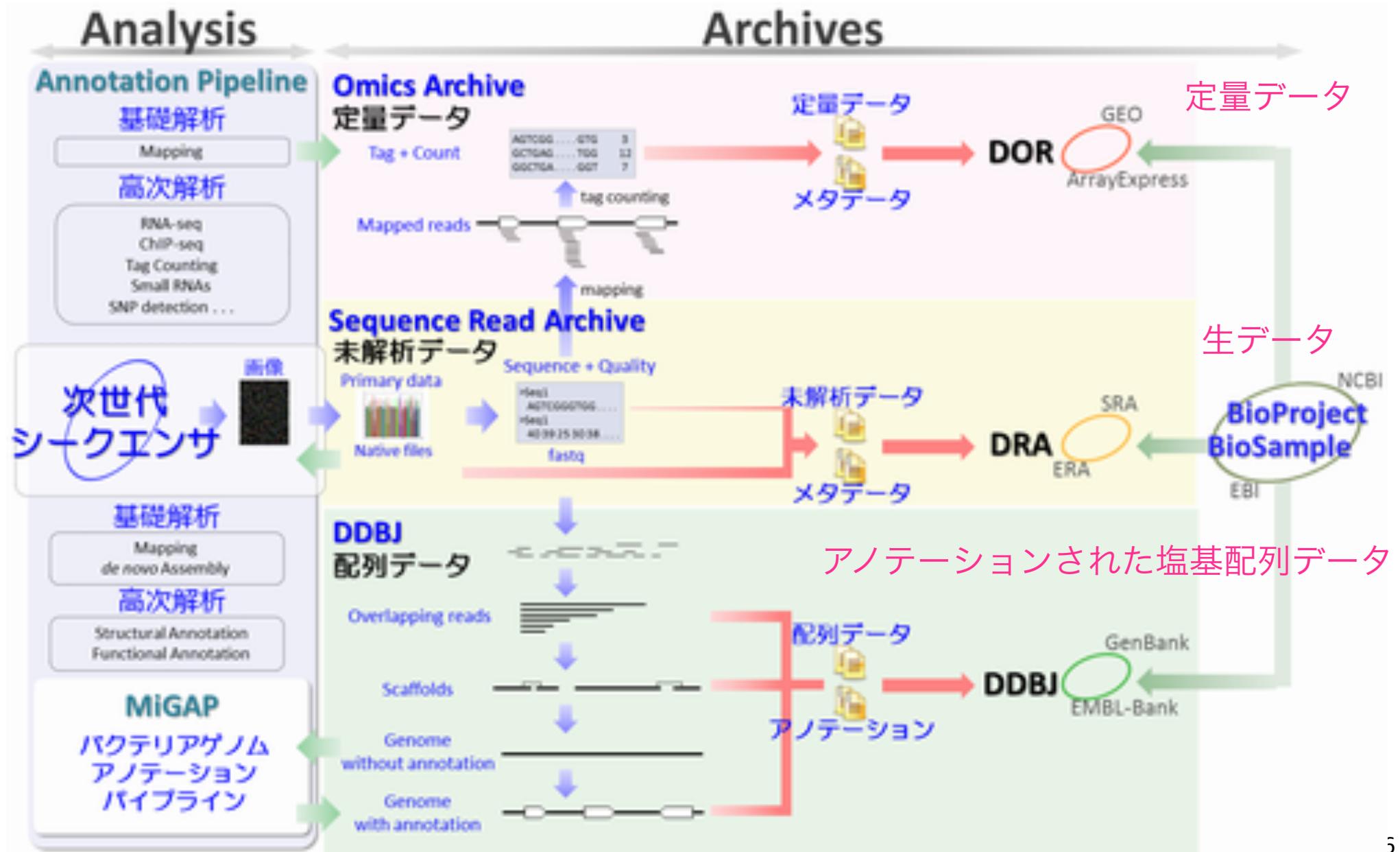
---

## ◆ SRA (NCBI)/ERA (EBI)/DRA (DDBJ)

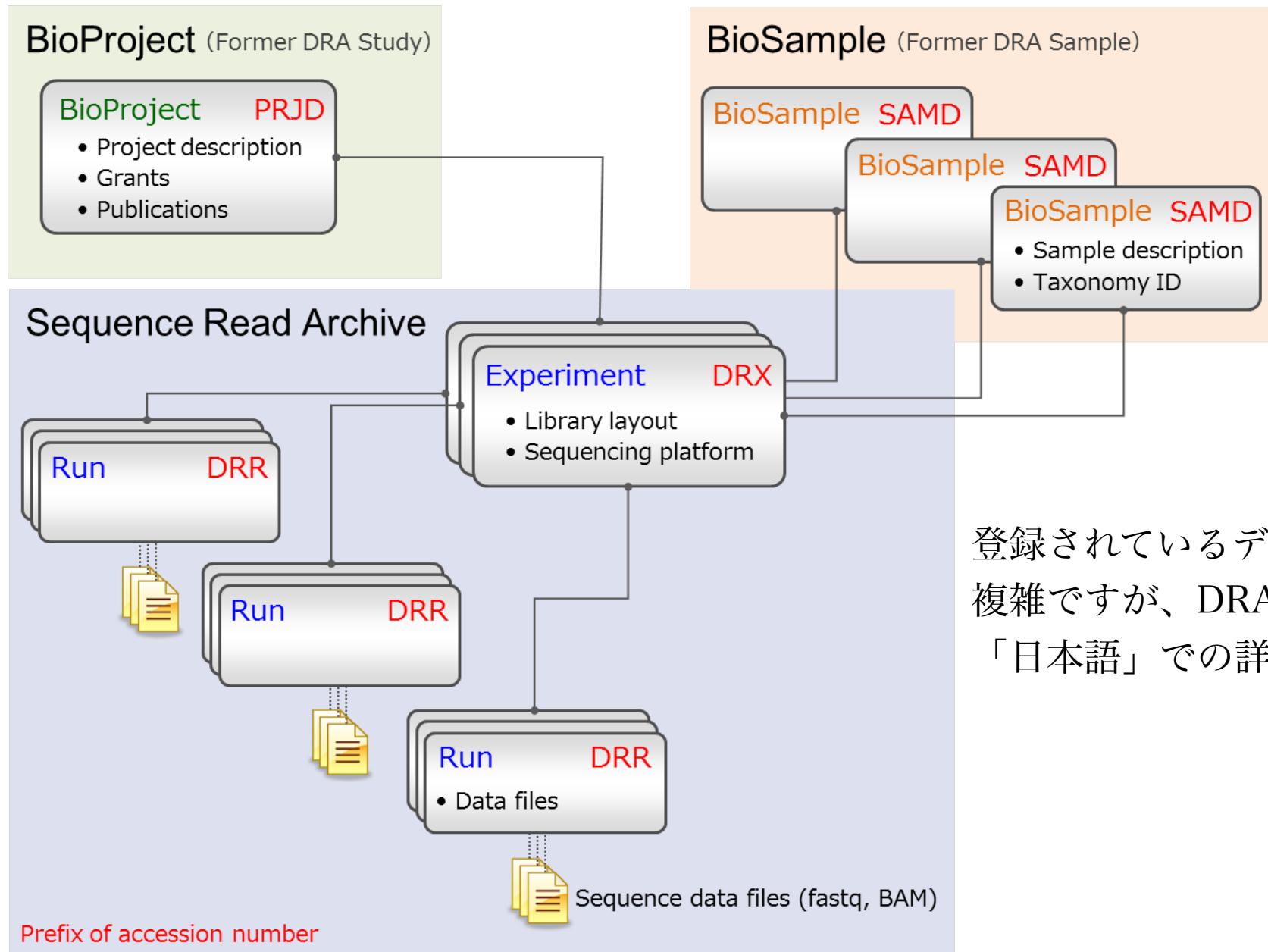
- 次世代シーケンスデータはどれかに登録する
  - 生データとマッピングデータ、メタデータのみ登録
  - 解析データは従来通りGenBank/ENA/DDBJに登録する
  - 定量データはGEO/ArrayExpress/DORに登録する
- 3極でデータを毎日交換しており、データの内容は同じ



# データの種類とデータベース



# 次世代シーケンスのデータ構造



# 実習

DRAsearchを使って  
次世代シーケンスデータを検索する

# DDBJ トップページ

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

The screenshot shows the DDBJ homepage with several red arrows highlighting specific features:

- An arrow points from the top navigation bar to the "DDBJ Service" section.
- Three arrows point to the "登録" (Registration) button, which is highlighted with a red circle.
- An arrow points from the "Hot Topics" section to the "登録をクリック" (Click Registration) text.

**DDBJ Service**

- 登録** (Data Submission)
- 検索・解析** (Search / Analysis)
- スパコン** (Super Computer)
- アーカイブ** (Archive)

**Hot Topics**

- 2014.06.27 PDB 2014-06-21 公開
- 2014.06.27 FAQ リニューアルのお知らせ
- 2014.06.27 「第29回 DDBJing 講習会 in 三島」の様子をYouTubeにアップしました
- 2014.06.26 Mass Submission System (MSS) 関連ツール更新のお知らせ
- 2014.06.19 DDBJ リリース 97.0, DAD リリース 67.0 完成
- 2014.06.19 PDB 2014-06-14 公開
- 2014.06.16 UniProt 2014\_06 公開

# DDBJデータ登録画面

塩基配列の登録 プロジェクトの登録 塩基配列登録の前に Flat File の説明 お問い合わせ

English

Google™カスタム検索 Search

HOME > データ登録

## データ登録

### データ登録方法

#### アノテーションをつけた塩基配列の登録

#### DDBJ Nucleotide Sequence Submission System

Web経由の塩基配列登録システム

#### Mass Submission System (MSS)

登録予定データが、件数が多い、多数の Feature を持つ、配列が長大、などの場合や、web 経由の登録システムが対応していないデータ（例：WGS）の登録

#### DDBJ Sequence Read Archive (DRA)

次世代シーケンサーから出力されたデータの登録

#### DDBJ Trace Archive (DTA)

Sanger法をベースとしたシーケンサーから得られるクロマトグラム (traces) の登録

#### DDBJ BioProject Database

研究プロジェクトの登録。データ(WGS, complete genome, transcriptome project, DRA, DTA など)をプロジェクト単位でグループ化するためのBioProject IDを発行。

#### DDBJ BioSample Database

実験データを得るために使用された生物学的な試料 (サンプル) についての情報を集中して管理するデータベース。BioSample IDを発行。

#### Japanese Genotype-phenotype Archive (JGA)

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の登録。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のみ登録可能。

### 登録データ種別

#### 塩基配列の登録・更新・修正

#### 登録前にお読みください

##### 塩基配列の登録について

登録に必要なデータや登録情報の具体的記述方法など

#### 修正・更新時にお読みください

##### 塩基配列登録データの修正・更新

#### その他のデータの登録・更新・修正

#### DRA/DTA/BioProject/BioSample ご利用の際にお読みください

##### D-way 登録アカウントマニュアル

データの登録に必要なD-wayアカウント取得の方法など

##### DRA/BioProject/BioSample データ登録マニュアル



# DRA ウェブサイト

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

Sequence Read Archive

Login & Submit | Databases ▾ | English | Contact

データ取得 Google™カスタム検索

Search

Home Handbook FAQ Search Download ▾ Pipeline About DRA

News 登録関連情報 データ検索 解析パイプライン

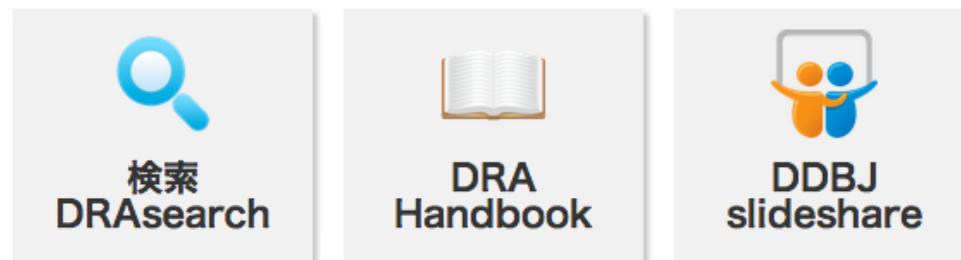
2014年05月13日: 新しい DRA 登録システムをリリース less...

新しい DRA 登録システムをリリースいたしました。主な変更点については変更点をまとめたスライド、新しいマニュアルをご覧ください。

(2014年6月6日)

2014年5月12日以前に作成されたステータスが new の登録の投稿作業を再開する場合、オブジェクトの追加・削除でエラーが発生する可能性があります。メタデータの内容をタブ区切りテキストファイルとしてダウンロードし、新しく作成した登録に内容をコピーし、投稿作業を再開することを推奨いたします。

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。



# データセットの検索：カテゴリを指定

DRA Search

Accession :

Organism :  StudyType :

CenterName :  Platform :

Keyword :

Show 20

Data Last Update 2014-07-01  
WebSite Last Update 2014-01-22

## Statistics

Released Entries

Type	Count
Submission	258755
Study	38619
Experiment	714047
Sample	688775
Run	831192

生物種の指定

(数文字入力すると候補が表示される)

## Organism

#	Organism Name	Study
1	Homo sapiens	2858
2	Mus musculus	2046
3	Drosophila melanogaster	649
4	unidentified	620
5	Caenorhabditis elegans	407
6	Arabidopsis thaliana	360
7	soil metagenome	342
8	marine metagenome	287
9	Saccharomyces cerevisiae	281
10	Solanum lycopersicum	187

## Study Type

#	Study Type	Study
1	Whole Genome Sequencing	21417
2	Other	6387
3	Transcriptome Analysis	4541
4	Metagenomics	2656
5	Epigenetics	1528
6		1440
7	Population Genomics	465
8	Exome Sequencing	89
9	Cancer Genomics	50
10	Pooled Clone Sequencing	31

## Center Name [All List]

#	Center Name	Study
1	BioProject	16234
2	GEO	5291
3	JGI	2665
4	UMIGS	2575
5		1440
6	JCVI	1161
7	BI	962
8	SC	790
9	WUSTL	535
10	WUGSC	506

StudyTypeを指定

Platformを指定

メタデータによる分類

# データセットの検索：キーワード検索

DRA Search

Accession :

Organism :  StudyType :

CenterName :  Platform :

Keyword :  stap

Show 20 records Search Study Search Clear

Data Last Update 2014-07-01  
WebSite Last Update 2014-01-22

## Statistics

Released Entries

Type	Count
Submission	258755
Study	38619
Experiment	714047
Sample	688775
Run	831192

何か適当なキーワードで検索

Organism

#	Organism Name	Study
1	<a href="#">Homo sapiens</a>	2858
2	<a href="#">Mus musculus</a>	2046
3	<a href="#">Drosophila melanogaster</a>	649
4	<a href="#">unidentified</a>	620
5	<a href="#">Caenorhabditis elegans</a>	407
6	<a href="#">Arabidopsis thaliana</a>	360
7	<a href="#">soil metagenome</a>	342
8	<a href="#">marine metagenome</a>	287
9	<a href="#">Saccharomyces cerevisiae</a>	281
10	<a href="#">Solanum lycopersicum</a>	187

Study Type

#	Study Type	Study
1	<a href="#">Whole Genome Sequencing</a>	21417
2	<a href="#">Other</a>	6387
3	<a href="#">Transcriptome Analysis</a>	4541
4	<a href="#">Metagenomics</a>	2656
5	<a href="#">Epigenetics</a>	1528
6		1440
7	<a href="#">Population Genomics</a>	465
8	<a href="#">Exome Sequencing</a>	89
9	<a href="#">Cancer Genomics</a>	50
10	<a href="#">Pooled Clone Sequencing</a>	31

Center Name [All List]

#	Center Name	Study
1	<a href="#">BioProject</a>	16234
2	<a href="#">GEO</a>	5291
3	<a href="#">JGI</a>	2665
4	<a href="#">UMIGS</a>	2575
5		1440
6	<a href="#">JCVI</a>	1161
7	<a href="#">BI</a>	962
8	<a href="#">SC</a>	790
9	<a href="#">WUSTL</a>	535
10	<a href="#">WUGSC</a>	506

メタデータによる分類

# DRA検索結果

**DRA Search**  Send Feedback [Search Home](#) [DRA Home](#)

Accession :

Organism :  StudyType :

CenterName :  Platform :

Keyword :

Show 20 records Sort by Study

**Search Results ( 4 records )** << < 1 / 1 Page > >>

**Filtered by**  
document type:experiment(1) run(1) sample(1) study(1)  
organism:Mus musculus(3)

#	META_FILE	ACCESSION	STUDY	STUDY_TITLE	STUDY_TYPE	ORGANISM	BASES	SUBMITTED	CENTER_NAME
1	<a href="#">SRA110029.run.xml</a> <a href="#">_ID&gt; &lt;/IDENTIFIERS&gt;</a> <a href="#">&lt;/EXPERIMENT_REF&gt;</a> <a href="#">&lt;/RUN&gt; &lt;RUN</a> alias="SMARTer <b>STAP</b> RNA-seq" accession="SRR1171578	<a href="#">SRR1171553</a> <a href="#">SRR1171554</a> <a href="#">SRR1171555</a> <a href="#">SRR1171556</a> <a href="#">SRR1171557</a> <a href="#">SRR1171558</a> <a href="#">SRR1171559</a> <a href="#">SRR1171560</a> <a href="#">SRR1171561</a> <a href="#">SRR1171562</a> <a href="#">SRR1171563</a> <a href="#">SRR1171564</a> <a href="#">SRR1171565</a> <a href="#">SRR1171566</a> <a href="#">SRR1171567</a> <a href="#">SRR1171568</a> <a href="#">SRR1171569</a> <a href="#">SRR1171570</a> <a href="#">SRR1171571</a> <a href="#">SRR1171572</a> <a href="#">SRR1171573</a> <a href="#">SRR1171574</a> <a href="#">SRR1171575</a> <a href="#">SRR1171576</a> <a href="#">SRR1171577</a> <a href="#">SRR1171578</a> <a href="#">SRR1171579</a> <a href="#">SRR1171580</a> <a href="#">SRR1171581</a> <a href="#">SRR1171582</a>	<a href="#">SRP038104</a>	<a href="#">Mus musculus</a> 1) Transcriptome or Gene expression; 2) histone modification H3K4me3, H3K27me3	<a href="#">Transcriptome Analysis</a>	<a href="#">Mus musculus</a>	182.6G	<a href="#">BioProject</a>	

BioProjectのIDをクリック

# BioProjectエントリ

DRA Search

Send Feedback

 Search Home

 DRA Home

SRP038104

## Study Detail

Title	Mus musculus 1)Transcriptome or Gene expression; 2)Histone Modification H3K4me3, H3K27me3
Study Type	Transcriptome Analysis
Abstract	Global Expression Profile and Epigenetic profile of STAP and other types of ES cells.
Description	
Center Name	BioProject

ExperimentのIDをクリック

## Navigation

 Submission	<a href="#">SRA110029</a>	
	<a href="#">SRA140515</a>	
 Experiment	<a href="#">SRX472627</a>	 
	<a href="#">SRX472628</a>	 
	<a href="#">SRX472629</a>	 
	<a href="#">SRX472630</a>	 
	<a href="#">SRX472631</a>	 
	<a href="#">SRX472632</a>	 
	<a href="#">SRX472633</a>	 
	<a href="#">SRX472634</a>	 
	<a href="#">SRX472635</a>	 
	<a href="#">SRX472636</a>	 
	<a href="#">SRX472637</a>	 
	<a href="#">SRX472638</a>	 
	<a href="#">SRX472639</a>	 

 Sample	<a href="#">SRS559080</a>
	<a href="#">SRS559081</a>
	<a href="#">SRS559082</a>
	<a href="#">SRS559083</a>

# Experimentエントリ

DRA Search

Send Feedback

 Search Home

 DRA Home

SRX472627  FASTQ  SRA

## Experiment Detail

Title	CD45 positive Cells ; RNASeq_Rep1
Design Description	TruSeq RNA Sample Prep Kit v2
Organism	Mus musculus

## Library Description

Name	
Strategy	RNA-Seq
Source	TRANSCRIPTOMIC
Selection	cDNA
Layout	PAIRED
Orientation	
Nominal Length	
Nominal Sdev	
Construction Protocol	

## Platform

Platform	ILLUMINA
Instrument Model	Illumina HiSeq 1500

## Processing

### Base Calls

### Sequence Space

### Base Caller

## Navigation

 Submission	<a href="#">SRA110029</a>	
 Study	<a href="#">SRP038104</a>	
 Sample	<a href="#">SRS559080</a>	
 Run	<a href="#">SRR1171553</a>	 

サンプル情報

データをダウンロード

# BioSampleエントリ



Send Feedback

Search Home

DRA Home

SRS559080

## Sample Detail

Title Generic sample from *Mus musculus*

Description

## Organism Info

Taxon ID [10090](#)

Common Name

Scientific Name *Mus musculus*

Anonymized Name

Individual Name

## Navigation

Submission [SRA110029](#) FTP

Study [SRP038104](#)

Experiment [SRX472627](#) FASTQ SRA

[SRX472628](#) FASTQ SRA

# DDBJ トップページ

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

The screenshot shows the DDBJ homepage with several key features highlighted:

- Header:** "DDBJ DNA Data Bank of Japan" logo, "English" button, and a search bar with "Google™ カスタム検索" and "Search" buttons.
- Main Navigation:** "DDBJ の紹介", "利用の手引き", "レポート・統計", "FAQ" (highlighted by a red arrow), and "お問い合わせ".
- Social Links:** RSS feed and DDBJ Twitter feed.
- International Collaboration:** Logos for DDBJ, INSDC, NCBI, and ENA/EBI, part of the "International Nucleotide Sequence Database Collaboration".
- DDBJ Service:** A section featuring four icons: "登録 Data Submission" (database and globe), "検索・解析 Search / Analysis" (magnifying glass over a DNA helix), "スパコン Super Computer" (two server racks), and "アーカイブ ftp. ddbj.nig.ac.jp" (upward arrow).
- Hot Topics:** A list of recent news items:

  - 2014.06.27 PDB 2014-06-21 公開
  - 2014.06.27 FAQ リニューアルのお知らせ
  - 2014.06.27 「第29回 DDBJing 講習会 in 三島」の様子をYouTubeにアップしました
  - 2014.06.26 Mass Submission System (MSS) 関連ツール更新のお知らせ
  - 2014.06.19 DDBJ リリース 97.0, DAD リリース 67.0 完成
  - 2014.06.19 PDB 2014-06-14 公開
  - 2014.06.16 UniProt 2014\_06 公開

- Bottom Footer:** CC-BY license logo and copyright notice: "©2014 統合データベース講習会 Licensed Under CC 表示 2.1".

# 遺伝研スーパーコンピュータ

---

## ◆ 計算ノード

- 64GB memory x 554 nodes
- 2TB memory x 10 nodes
- 10TB memory x 1 node

## ◆ Storage

- 7 PB 高速HDD
- 5.5 PB 大容量省電力HDD

## ◆ 利用申請することで無料で利用できます！

- コマンドラインが基本
- (停電で結構頻繁に落ちる)

# スパコン利用申請はこちら

**NIG**  
SUPERCOMPUTER

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

現在地: Home 検索... 検索 サイトポリシー サイトマップ

2014年07月01日

重要なお知らせ

一覧へ

公開日	表題
2014年6月30日	【ssh接続障害】Phase1システムゲートウェイ障害のお詫び
2014年6月27日	【計画停止:4】GPU計算ノードアップデートのお知らせ
2014年6月26日	【計画停止:3】Fat計算ノードサービス再開のお知らせ
2014年6月25日	【計画停止:2】スパコンサービス再開のお知らせ
2014年3月4日	2014年3月5日からのスパコンPhase2システムご利用方法について

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

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



システム構成

- ハードウェア構成
- ソフトウェア構成
- プログラミング環境
- 利用可能バイオツール
- 利用可能OSS
- 利用可能DB

システム使用方法

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

ディスク利用状況

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

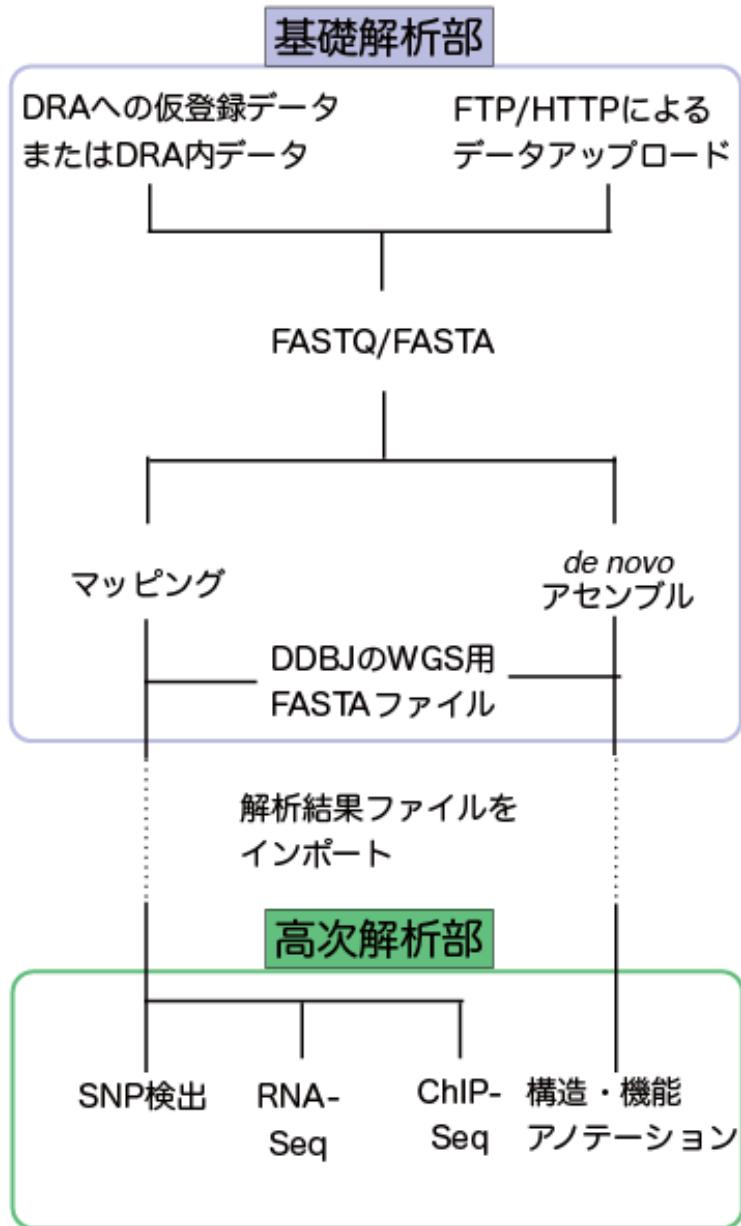
ディスク	size	quota	状況
lustre1	980.81 / 1,024.90	95%	lustre1 file : 144,176,194 / 731,840,512 19%
quota	2,037.73 / 1,024.90	198%	
lustre2	970.45 / 1,024.90	94%	lustre2 file : 525,058,984 / 731,840,512 71%
quota	997.45 / 1,024.90	97%	
lustre3	178.50 / 1,787.70	9%	lustre3 file : 14,888,401 / 731,840,512 2%
quota			

各種申請

- 各種申請窓口について
- 新規ユーザ登録申請
- 大規模利用申請
- MiGAP利用申請
- DDBJ PipeLine利用申請
- OSSインストール申請
- アカウント継続・停止申請

(cc) BY ©2014 統合データベース講習会 Licensed Under CC 表示 2.1

# DDBJ解析パイプライン



## DDBJ Read Annotation Pipeline

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

Galaxy / P-GALAXY

<https://p-galaxy.ddbj.nig.ac.jp/>

DDBJのスパコンをウェブ経由で利用

# DDBJ Read Annotation Pipeline

11種類のマッピング・アセンブルソフトに対応

## マッピング系

BLAT	高速シーケンサ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したがい開発は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	メモリ消費量が少なく、より高速。精度はBWAより若干落ちる
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に対してインデックスを作成、高速でマッピングする
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

## アセンブリ系

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リ
Velvet	高速シーケンサ登場初期に開発された。メモリ消費多め
Trinity	RNA-Seq配列のアセンブリ。上記3つともにde bruijn graphというアルゴリズムを使用

# 実習

DDBJ Read Annotation Pipeline  
を使って次世代シーケンスのデータを解析する

# DRA ウェブサイト

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

The screenshot shows the DDBJ Sequence Read Archive (DRA) website. At the top left is the DDBJ logo. To its right are links for 'Login & Submit', 'Databases ▾', 'English', and 'Contact'. Below these is a search bar containing 'Google™ フルテキスト検索' and a magnifying glass icon. The main title 'Sequence Read Archive' is centered above a navigation bar with links for 'Home', 'Handbook', 'FAQ', 'Search', 'Download ▾', 'Pipeline', and 'About DRA'. A red triangle icon is positioned above the 'About DRA' link.

## News

### 解析パイプライン

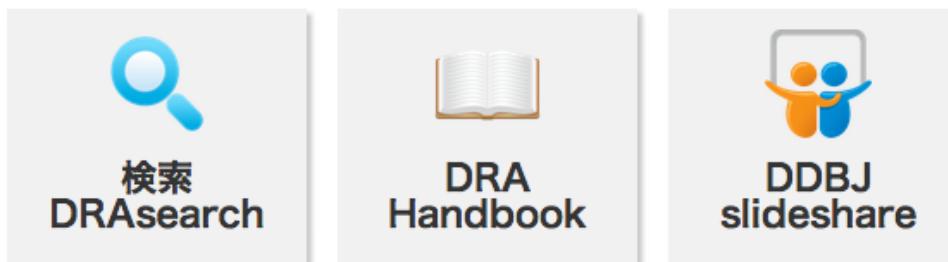
2014年05月13日: 新しい DRA 登録システムをリリース [less...](#)

新しい DRA 登録システムをリリースいたしました。主な変更点については変更点をまとめたスライド、新しいマニュアルをご覧ください。

(2014年6月6日)

2014年5月12日以前に作成されたステータスが new の登録の投稿作業を再開する場合、オブジェクトの追加・削除でエラーが発生する可能性があります。メタデータの内容をタブ区切りテキストファイルとしてダウンロードし、新しく作成した登録に内容をコピーし、投稿作業を再開することを推奨いたします。

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。



©2014 統合データベース講習会 Licensed Under CC 表示 2.1

# DDBJ Read Annotation Pipeline



## 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。

**LOG IN**

**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 TD; User[ユーザ] -- "File Upload" --> Reads[Reads, metadata]; Reads --> DRA[DRA Start]; DRA --> Archive[DDBJ Read Archive]; Archive --> WGS[WGS/CON]; WGS --> Manual[Manual curation]; Manual --> Annotation[metadata, annotation]; Annotation --> DDBJ[DDBJ]; DDBJ --> Position[map positions]; Position --> Foundation[基礎処理部]; Foundation --> Mapping[Mapping]; Foundation --> Assembly[de novo assembly]; Mapping --> Foundation; Assembly --> Foundation; Foundation --> Workflow[解析目的別ワークフロー]; Workflow --> Genome[ゲノム解析]; Workflow --> RNAseq[RNA-seq]; Workflow --> ChIPseq[ChIP-seq];
```

**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](#)
- [FAQ : FTPでファイルをアップロードしDDBJ Pipelineへ登録する方法](#)
- [チュートリアル : DDBJ PipelineでHGAP法でPacBioリードのアンブリを行う方法](#)

**DRAアカウント登録**  
DRAアカウントの登録に関しては [こちらをご覧ください。](#)

**Tweets** [Follow](#)

pipeline @pipeline\_info 11 Jun  
!!!!Notice!!!! DDBJ pipeline services will NOT be available due to system maintenance as follows.  
2014/6/23(Mon)13:00 - 6/26(Thu)12:00 (JST)

# データセットの選択

FTPで手持ちのデータを  
アップロード

DDBJ  
DNA Data Bank of Japan

**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.

DRAからデータをインポート

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 : DRA000001

TYPE	ACCESSION	ALIAS	FILENAME	DL	VIEW
Submission	DRA000001	DRA000001	DRA000001.submission.xml	<input type="button" value="DownLoad"/>	<input type="button" value="View"/>
Sample	DRS000001	DRS000001	DRA000001.sample.xml	<input type="button" value="DownLoad"/>	<input type="button" value="View"/>
Study	DRP000001	DRP000001	DRA000001.study.xml	<input type="button" value="DownLoad"/>	<input type="button" value="View"/>
Experiment	DRX000001	DRX000001	DRA000001.experiment.xml	<input type="button" value="DownLoad"/>	<input type="button" value="View"/>
Run	DRR000001	DRR000001	DRA000001.run.xml	<input type="button" value="DownLoad"/>	<input type="button" value="View"/>

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.

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.)

HTTPで手持ちのデータを  
アップロード

# SRA/DRAデータのインポート

## 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

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 e  
When the status makes "failed" or "preparing", please retry it.

queued : waiting or during download, done : file is ready, failed  
DRA uncheckd : download is ok, but md5 was not check.

Status	Submission	Request
done	SRA009211	2012-0
preparing	SRA026538	2012-0

### Confirmation

Click a OK button to start import.

This operation may take several minutes to several hours.

#### Option

Send a mail when completed importing.

Show a accessions list.

※実行しないでください！

OK

Cancel

# インポートされたデータ

DDBJ  
DNA Data Bank of Japan

ACCOUNT  
login ID [orenoddb]  
Logout  
Change password

ANALYSIS  
Data setup  
DRA Start  
FTP upload  
HTTP upload  
DRA Import  
Preprocessing Start  
step 1 Preprocessing 1  
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 SRA009211 2

TYPE	ACCESSION	ALIAS	FILENAME	DL
Submission	SRA009211	VitisSolexa	SRA009211.submission.xml	<a href="#">DownLoad</a> <a href="#">View</a>
Sample	SRS008099 SRS008102 SRS008103 SRS008104 SRS008105 SRS008106 SRS008107 SRS008108 SRS008109 SRS008110 SRS008111 SRS008112 SRS008113 SRS008114 SRS008115 SRS008116 SRS008117	muscat_s0l01 thompson_s0l02 colombard_s0l03 gewurztraminer_s0l04 kadarka_s0l05 malvasia_s0l06 pinotnoir_s0l07 plavacmaill_s0l08 vsylvestris_s0l09 vamurensis_s0l10 vcinerea_s0l11 vlabrusca_s0l12 vpaimata_s0l13 vrondifolia_s0l16 riesling_s0l14 ehrenfeiser_s0l15 inbred_pinotnoir	SRA009211.sample.xml	<a href="#">DownLo</a>
Study	SRP001055	Rapid Genomic Characterization of the Genus Vitis	SRA009211.study.xml	<a href="#">DownLo</a>
Experiment	SRX013989 SRX013990 SRX013991 SRX013992 SRX013993 SRX013994 SRX013995 SRX013996 SRX013997 SRX013998 SRX013999	Muscat RRL thompson RRL colombard RRL Gewurztraminer RRL Kadarka RRL Malvasia RRL pinotnoir RRL plavacmaill RRL Vsylvestris RRL Vamurensis RRL Vcinerea RRL	SRA009211.experiment.xml	<a href="#">DownLo</a>

STUDY TITLE Rapid Genomic Characterization of the Genus Vitis  
STUDY TYPE Population Genomics

Select your registered query files.

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	SRX013989	SRS008099	SRR031097					ILLUMINA	single
2	SRX013990	SRS008102	SRR031098					ILLUMINA	single
3	SRX013991	SRS008103	SRR031099					ILLUMINA	single
4	SRX013992	SRS008104	SRR031100					ILLUMINA	single
5	SRX013993	SRS008105	SRR031101					ILLUMINA	single
6	SRX013995	SRS008107	SRR031102					ILLUMINA	single
7	SRX013996	SRS008108	SRR031103					ILLUMINA	single
8	SRX013997	SRS008109	SRR031104					ILLUMINA	single
9	SRX013998	SRS008110	SRR031105					ILLUMINA	single
10	SRX013999	SRS008111	SRR031106					ILLUMINA	single
11	SRX014000	SRS008112	SRR031107					ILLUMINA	single
12	SRX014001	SRS008113	SRR031108					ILLUMINA	single
13	SRX014002	SRS008115	SRR031109					ILLUMINA	single
14	SRX014003	SRS008116	SRR031110					ILLUMINA	single
15	SRX014004	SRS008114	SRR031111					ILLUMINA	single
16	SRX014005	SRS008117	SRR031112					ILLUMINA	single
17	SRX013994	SRS008106	SRR031124					ILLUMINA	single

: from metadata : Counted from FASTQ (Sequence length is calculated from the first entry.)

3 4

Preprocessing -> 計算対象ランデータを選択 -> NEXT

NEXT

# Preprocessing (クオリティチェック)

Set Parameters for Preprocessing

Your selected queries

Run ACCESSION	Read length	Quality Score	Read Layout
SRR031112 ->	bp	single	

Steps of preprocessing workflow

Step1: Set the encoding type of the quality values for sequence.

Phred+33  Phred+64  
If you don't know it, please see '2.2 Encoding' of this site.

Step2: BASE TRIMMING with low quality from 5'end and 3'end of each read.

Bases with low quality ( $QV \leq \text{THRESHOLD}$ ) are trimmed from 5'end and 3'end of each read. Bases of the trimmed read indicate high quality ( $QV > \text{THRESHOLD}$ ). If read length after base trimming is too short (length  $\leq 24$  bp), the read is removed. Thus length will be 25bp.

QV THRESHOLD: 19

Step3: READ REMOVING to discard trimmed reads including low quality bases with high percentage.

Trimmed reads with high percentage ( $\geq \text{Low quality bases} / \text{Total bases}$ ) of the low quality bases ( $QV \leq \text{THRESHOLD}$ ) are discarded.

QV THRESHOLD: 14  
Percentage THRESHOLD: 30

Step 4: In the case of paired-end read, the pair is discarded when one read of the pair is removed at 'Step2' or 'Step3'.

Run Confirmation

Email notification

Send email notification when the job is completed or aborted with error.  
kawano@dbcls.jp \* Required

Confirmation of entries

Query sets

SRR031112 - inbred\_pinotnoir

BACK RUN

確認画面

1

2

3

パラメータ設定

BACK NEXT

※ログインしないと実行できません (Runボタンがでてきません)

# 進捗確認画面-Preprocessing

**Status - Preprocessing**

		Mapping Job	de novo Assembly Job					
Order								
Sort by : ID Descending <input type="checkbox"/> Show Only Your Own Job Reload								
<input type="button" value="Delete (N.A.)"/>								
ID	User ID	Files	P/S	Status	Read #	Read length	Detail	Start time End time
4166	---	Preprocess_Sai	P	complete		---		2012-07-31 14:28:19
								2012-07-31 14:28:56
4165	---	e_coli_MAPPIN	P	complete		---		2012-07-31 14:28:19
								2012-07-31 14:28:51
4098	---	SRA029956 737814100930.	S	complete		---		2012-07-25 18:23:43
								2012-07-25 18:24:24
4095	orenoddbj	SRA009211 inbred_pinotnoi	S	complete		---	<input type="button" value="View"/>	2012-07-25 16:56:32
								2012-07-25 17:04:58
4094	orenoddbj	SRA009211 muscat_sol01	S	complete		---	<input type="button" value="View"/>	2012-07-25 15:37:42
								2012-07-25 15:44:04
4093	orenoddbj	SRA009211 muscat_sol01	S	complete		---	<input type="button" value="View"/>	2012-07-25 15:37:21
								2012-07-25 15:43:50

**Detail view**

**Job info**

ID 4095	Tool (Version) (1.0)
RunAccession or Filename SRR031112	Download <a href="#">SRR031112.fastq.bz2</a>
File SRR031112.fastq.bz2	Fastq Download <a href="#">download (541.1 MB)</a>
<a href="#">QS Average (PDF)</a> <a href="#">QS Count (PDF)</a> <a href="#">QS Error (PDF)</a>	
<a href="#">download (6.7 KB)</a> <a href="#">download (5.2 KB)</a> <a href="#">download (5.3 KB)</a>	

**Time**

Wait time	Start time	End time
0: 2:36	2012-07-25 16:56:32	2012-07-25 17:04:58

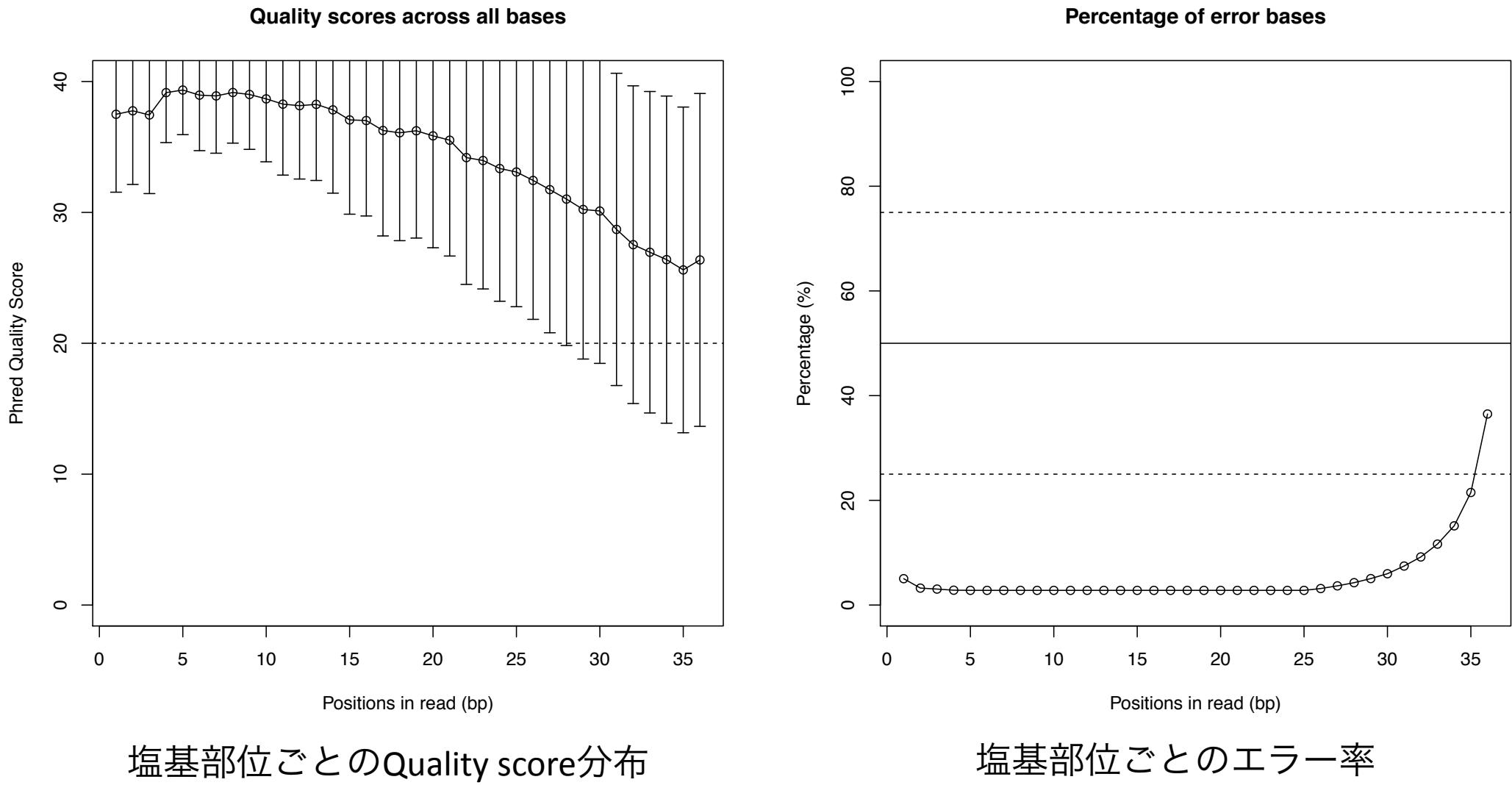
Command	Start time	End time	Log1	Log2	Result	MD5
perl avg_p.pl fqlist.txt qscore	2012-07-25 16:56:33	2012-07-25 17:01:09	<a href="#">View</a>			
perl pdel_p3_tpl fqlist.txt qscore20 25 00 100	2012-07-25 17:01:09	2012-07-25 17:04:45	<a href="#">View</a>			
perl user_fastq_copy.pl preprocessing.xml orenoddbj	2012-07-25 17:04:46	2012-07-25 17:04:58		<a href="#">View</a>		

**結果**

背景無色（白）が他のユーザ  
背景黄色が自分のジョブ

**BACK**

# Preprocessing 結果



塩基部位ごとのQuality score分布

塩基部位ごとのエラー率

どの塩基まで解析に使用するかの目安になる

# パイプラインの実行～マッピング

DDBJ  
DNA Data Bank of Japan

ACCOUNT  
login ID [orenoddb]  
Logout  
Change password

ANALYSIS  
Data entry 1  
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: SRA009211 2

TYPE	ACCESSION	ALIAS	FILENAME	DL	VIEW
Submission	SRA009211	VitisSolexa	SRA009211.submission.xml	DownLoad	View
Sample	SRS008099 SRS008102 SRS008103 SRS008104 SRS008105 SRS008106 SRS008107 SRS008108 SRS008109 SRS008110 SRS008111 SRS008112 SRS008113 SRS008114 SRS008115 SRS008116 SRS008117	muscat_sol01 thompson_sol02 colombard_sol03 gewurztraminer_sol04 kadarka_sol05 malvasia_sol06 pinotnoir_sol07 plavacmaill_sol08 vsilvestris_sol09 vamurensis_sol10 vcinerea_sol11 vlabrusca_sol12 vpaimata_sol13 vrundifolia_sol16 riesling_sol14 ehrenfeiser_sol15 inbred_pinotnoir	SRA009211.sample.xml	DownLo	
Study	SRP001055	Rapid Genomic Characterization of the Genus Vitis	SRA009211.study.xml	DownLo	
Experiment	SRX013989 SRX013990 SRX013991 SRX013992 SRX013993 SRX013994 SRX013995 SRX013996 SRX013997 SRX013998 SRX013999	Muscot RRL thompson RRL colombard RRL Gewurztraminer RRL Kadarka RRL Malvasia RRL pinotnoir RRL plavacmaill RRL Vsylvestris RRL Vamurensis RRL Vcinerea RRL	SRA009211.experiment.xml	DownLo	

STUDY TITLE: Rapid Genomic Characterization of the Genus Vitis  
STUDY TYPE: Population Genomics

Select your registered query files.

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	SRX013989	SRS008099	SRR031097					ILLUMINA	single
2	SRX013990	SRS008102	SRR031098					ILLUMINA	single
3	SRX013991	SRS008103	SRR031099					ILLUMINA	single
4	SRX013992	SRS008104	SRR031100					ILLUMINA	single
5	SRX013993	SRS008105	SRR031101					ILLUMINA	single
6	SRX013995	SRS008107	SRR031102					ILLUMINA	single
7	SRX013996	SRS008108	SRR031103					ILLUMINA	single
8	SRX013997	SRS008109	SRR031104					ILLUMINA	single
9	SRX013998	SRS008110	SRR031105					ILLUMINA	single
10	SRX013999	SRS008111	SRR031106					ILLUMINA	single
11	SRX014000	SRS008112	SRR031107					ILLUMINA	single
12	SRX014001	SRS008113	SRR031108					ILLUMINA	single
13	SRX014002	SRS008115	SRR031109					ILLUMINA	single
14	SRX014003	SRS008116	SRR031110					ILLUMINA	single
15	SRX014004	SRS008114	SRR031111					ILLUMINA	single
16	SRX014005	SRS008117	SRR031112					ILLUMINA	single
17	SRX013994	SRS008106	SRR031124					ILLUMINA	single

: from metadata : Counted from FASTQ (Sequence length is calculated from the first entry.)

3

4

DRA Startで初期画面に戻る

計算対象ランデータを選択 -> NEXT

NEXT

# 解析ツールの選択

マッピング系ツール

アセンブリ系ツール

Selecting Tools for Basic Analysis of DDBJ ANNOTATION PIPELINE

BACK NEXT

Reference Genome Mapping

Tool	Help	Version	Input data			Evaluation			Analysis		Output format		Comment
			Base space	Color space	Paired end	Depth	Coverage	Error rate	SNP	Indel	.gff	.bed	
<input type="checkbox"/> BLAT		34	✓										Single-end analysis only
<input type="checkbox"/> Mag		0.7.1	✓		✓				✓	✓	✓	✓	
<input checked="" 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 50 bp, Bowtie2 is generally faster, more sensitive, and uses less memory than Bowtie1.
<input type="checkbox"/> TopHat2		2.0.9	✓		✓			✓					

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 severe recommend when performing Velvet, total length of those reads is up to 22G bp. Maximum K-mer value is 64.
<input type="checkbox"/> Trinity		r2013-02-25			✓		RNA-Seq De novo Assembly
<input type="checkbox"/> Platanus		1.2.2			✓		
<input type="checkbox"/> HGAP		Protocol3 (v 2.2.0)					HGAP Pipeline for PacBio Sequence based on SMRT Analysis v2.2.0. For bax.h5 file only. (Beta version)

Mapping Contigs by de novo Assemble to Reference Sequences.  
The contigs will be aligned to reference genome.

Tool	Comment
<input checked="" type="radio"/> BLAT	Single-end analysis only

# 解析するランの指定

Generating Query Sets from Query Read Files

RESET BACK NEXT

Single analysis  
Layout of single sequence.

5' 3'  
Linker(1) Target Linker(2)

	Run ACCESSION	Read length	Quality Score
1	<input checked="" type="checkbox"/> SRR031112 ->	bp	

2 confirm

QUERY SET

RESET BACK NEXT

Generating Query Sets from Query Read Files

RESET BACK NEXT

Single analysis  
Layout of single sequence.

5' 3'  
Linker(1) Target Linker(2)

	Run ACCESSION	Read length	Quality Score

confirm

QUERY SET  
Query set1

PairedOrientation	RunAccession	RunAlias	RowLength	QualityScore1	QualityScore2
single	SRR031112	inbred_pinotnoir			

3  
RESET BACK NEXT

# リファレンスゲノムの指定

Specifying Database of Reference Genome

RESET BACK NEXT

Major genome sets

Organisms: Arabidopsis thaliana  
Genome sets: TAIR8

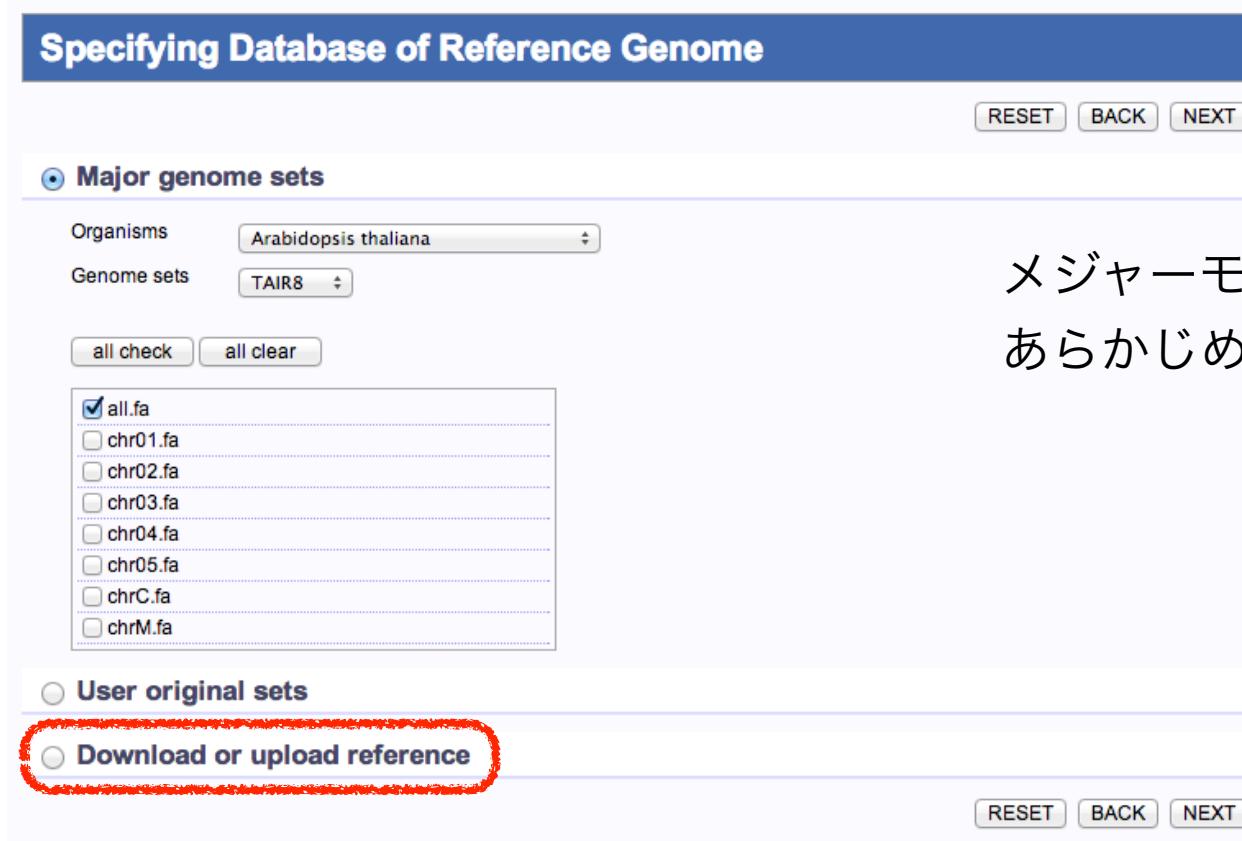
all check  all clear

all.fa  
 chr01.fa  
 chr02.fa  
 chr03.fa  
 chr04.fa  
 chr05.fa  
 chrC.fa  
 chrM.fa

User original sets

Download or upload reference

RESET BACK NEXT



メジャー モデル生物のリファレンスゲノムは  
あらかじめプリセットされている

ヒト  
マウス  
線虫  
イネ  
シロイヌナズナ  
酵母  
...

その他のリファレンスゲノムは

- 自分でゲノム配列をアップロードする
- DDBJからインポートする

# NCBIゲノムページ

NCBI Resources How To Sign in to NCBI

Genome Genome txid11886[Organism:noexp] Search Save search Limits Advanced Help

Display Settings: Overview Send to: Tools

Organism Overview ; Genome Assembly and Annotation report [1]

ID: 5120

## Rous sarcoma virus

Rous sarcoma virus RefSeq Genome

Lineage: Viruses[3915]; Retro-transcribing viruses[118]; Retroviridae[55]; Orthoretrovirinae[44]; Alpharetrovirus[7]; Rous sarcoma virus[1]

• Rous sarcoma virus

Submitter: NLM, NIH, USA, Bethesda

Status: Complete

BioProjects: PRJNA14978

Type	Name	RefSeq	INSDC	Size (Kb)	GC%	Protein	Gene
	-	NC_001407.1	AF033808.1	9.39	54.1	4	4

Genome Region

Go to nucleotide Graphics FASTA GenBank

Search See more...

txid11886[Organism:noexp]

# リファレンスゲノムデータのインポート

Specifying Database of Reference Genome

RESET BACK NEXT

Major genome sets  
 User original sets  
 Download or upload reference

Retrieving a chromosome from DDBJ-DB by using HTTP REST

Input Accession Number (INSD) or (RefseqID)  
FN597015  
LOAD

INTERNET PIPELINE DDBJ-DB

Request  
HTTP REST \* Data (fasta)  
\* Representational State Transfer(REST)

Uploading reference from local drive.  
FASTA only ファイルを選択 選択されていません  
2GB Filesize Limit

UPLOAD

RESET BACK NEXT

Vitis vinifera 各クロモソームのINSD IDを  
→ 入力、LOAD を繰り返す  
→ CREATE DATASET  
→ CREATE GENOMESET  
でリファレンスゲノムを登録する

>FN597015 FN597015.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 1, chr1.	DELETE
>FN597017 FN597017.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 2, chr2.	DELETE
>FN597018 FN597018.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 3, chr3.	DELETE
>FN597020 FN597020.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 4, chr4.	DELETE
>FN597022 FN597022.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 5, chr5.	DELETE
>FN597024 FN597024.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, chromosome 6, chr6.	DELETE
>FN597025 FN597025.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 7, chr7.	DELETE
>FN597027 FN597027.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 8, chr8.	DELETE
>FN597028 FN597028.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 9, chr9.	DELETE
>FN597030 FN597030.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 10, chr10.	DELETE
>FN597032 FN597032.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 11, chr11.	DELETE
>FN597034 FN597034.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 12, chr12.	DELETE
>FN597036 FN597036.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 13, chr13.	DELETE
>FN597038 FN597038.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 14, chr14.	DELETE
>FN597039 FN597039.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 15, chr15.	DELETE
>FN597040 FN597040.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 16, chr16.	DELETE
>FN597042 FN597042.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 17, chr17.	DELETE
>FN597044 FN597044.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 18, chr18.	DELETE
>FN597046 FN597046.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 19, chr19.	DELETE

CREATE DATASET

77

# リファレンスゲノムの選択

**Specifying Database of Reference Genome**

RESET BACK NEXT

Major genome sets

User original sets

Genome sets

>FN597015|FN597015.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 1, chr1.

>FN597022|FN597022.1 Vitis vinifera, whole genome shotgun sequence of line PN40024  
>FN597024|FN597024.1 Vitis vinifera, whole genome shotgun sequence of line PN40024  
>FN597025|FN597025.1 Vitis vinifera, whole genome shotgun sequence of line PN40024  
>FN597027|FN597027.1 Vitis vinifera, whole genome shotgun sequence of line PN40024  
>FN597028|FN597028.1 Vitis vinifera, whole genome shotgun sequence of line PN40024  
>FN597030|FN597030.1 Vitis vinifera, whole genome shotgun sequence of line PN40024  
>FN597032|FN597032.1 Vitis vinifera, whole genome shotgun sequence of line PN40024  
>FN597034|FN597034.1 Vitis vinifera, whole genome shotgun sequence of line PN40024  
>FN597036|FN597036.1 Vitis vinifera, whole genome shotgun sequence of line PN40024  
>FN597038|FN597038.1 Vitis vinifera, whole genome shotgun sequence of line PN40024  
>FN597039|FN597039.1 Vitis vinifera, whole genome shotgun sequence of line PN40024  
>FN597040|FN597040.1 Vitis vinifera, whole genome shotgun sequence of line PN40024  
>FN597042|FN597042.1 Vitis vinifera, whole genome shotgun sequence of line PN40024  
>FN597044|FN597044.1 Vitis vinifera, whole genome shotgun sequence of line PN40024  
>FN597046|FN597046.1 Vitis vinifera, whole genome shotgun sequence of line PN40024  
>FN597015|FN597015.1 Vitis vinifera, whole genome shotgun sequence of line PN40024  
>FN597020|FN597020.1 Vitis vinifera, whole genome shotgun sequence of line PN40024  
>FN597018|FN597018.1 Vitis vinifera, whole genome shotgun sequence of line PN40024  
>FN597017|FN597017.1 Vitis vinifera, whole genome shotgun sequence of line PN40024  
>FN597015|FN597015.1 Vitis vinifera, whole genome shotgun sequence of line PN40024

Download or upload reference

RESET BACK NEXT

Vitis vinifera を選択

# パラメータ設定、確認、実行

## Setting for Reference Genome Mapping

**bwa**

### Set optional parameters of the single-end analysis

**Step1) Convert reference sequence**

bwa index -a is (for small-size reference) refgenome.fasta

[Options usage \(click\)](#)

**Step2) Map**

bwa aln -t 4 refgenome.fasta query.fastq(fasta) > out.sam

bwa samse (for short query) refgenome.fasta in.sam  
query1.fastq(fasta) > out.sam

[Options usage \(click\)](#)

**Step3)'uniq': Remove multiple hits on the genome from out.sam.**

Please choose uniq mode.

Do not remove any read.  
 Discard multiply mapped reads, and Retain uniquely mapped reads.

**Step4) Convert the read alignment to .BAM format**

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

**Step5) Detect DNA polymorphism**

Please choose one of the following.

samtools pileup -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.filter.vcf

**Step6) 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.

**Step7) Create assembled sequences in FASTA file from pileupped reads to submit WGS division of DDBJ.**

perl getConsGeno\_4pipeline.pl pileupFile Not to include insertion of pileupped reads. out\_WGS.txt

\* Threshold of insertion of pileupped reads: the quality threshold for indels <= 50 and allele constitutes 80% of pileupped reads.

[BACK](#) [NEXT](#)

## Run Confirmation

[BACK](#)

**Destination of mail**  
When the request is completed, the system sends an email to this address.  
 \* Required

Result files will be deleted 60 days after submission.

**Reference Genome Map [bwa]**

**Query sets**

Query set1	PairedOrientation	RunAccession	RunAlias	RowLength	QualityScore1	QualityScore2
single	SRR031112	inbred_pinotnoir				

**genome sets**

>FN597015|FN597015.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 1, chr1.  
• >FN597022|FN597022.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 5, chr5.  
• >FN597024|FN597024.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, chromosome 6, chr6.  
• >FN597025|FN597025.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 7, chr7.  
• >FN597027|FN597027.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 8, chr8.  
• >FN597028|FN597028.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 9, chr9.  
• >FN597030|FN597030.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 10, chr10.  
• >FN597032|FN597032.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 11, chr11.  
• >FN597034|FN597034.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 12, chr12.  
• >FN597036|FN597036.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 13, chr13.  
• >FN597038|FN597038.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 14, chr14.  
• >FN597039|FN597039.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 15, chr15.  
• >FN597040|FN597040.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 16, chr16.  
• >FN597042|FN597042.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 17, chr17.  
• >FN597044|FN597044.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 18, chr18.  
• >FN597046|FN597046.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 19, chr19.  
• >FN597015|FN597015.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 1, chr1.  
• >FN597020|FN597020.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 4, chr4.  
• >FN597018|FN597018.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 3, chr3.  
• >FN597017|FN597017.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 2, chr2.  
• >FN597015|FN597015.1 Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 1, chr1.

**Command Options**  
bwa

計算に時間がかかるのでメールアドレスの登録は必須  
ログインしないと実行できません  
(Runボタンが出てきません)

# 進捗確認画面-Mapping

DDBJ  
DNA Data Bank of Japan

**ACCOUNT**  
login ID [orenoddbj]  
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.  
step2.  
Mapping  
step1.  
de novo Assembly  
step2-All status

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 (N/A)

ID	User ID	Submission accession	P/S	Status	Tool	Read #	Read length	Genome size	Detail	Start time	End time	Elapsed time
4170	orenoddbj	SRA009211 inbred_pinotnoir	S	generating	bwa	---	---	228 M	<a href="#">View</a>	---	---	---
4168	--	---	P	running	bwa	122,403,348	---	---	2012-07-31 18:47:13	---	---	---
4167	--	---	P	complete	Bowtie	4,296,782	---	3 M	2012-07-31 15:11:54	01:12:54	16:24:08	
4164	--	rq10_2	P	running	bwa	4,296,782	---	3 M	2012-07-30 15:24:01	---	---	
4161	--	---	S	error	bwa	---	---	---	---	---	---	
4156	--	rq10_2	P	error	bwa	0	---	3 M	---	---	---	
4155	--	rq10	P	error	bwa	0	---	3 M	---	---	---	

背景無色（白）が他のユーザ  
背景黄色が自分のジョブ

**Detail view** BACK

**Job info**

ID	4170
Tool (Version)	bwa (0.5.9)
RunAccession or Filename	SRR031112
Download	<a href="#">SRR031112.fastq.bz2</a>
Read length	N.A. bp
Alias	inbred_pinotnoir

**Genome set**

---

**Chromosome**

FN597017\_120725164226153 FN597018\_120725164259490 FN597020\_120725164327389 FN597022\_120725164403873  
FN597024\_120725164450583 FN597025\_120725164519672 FN597027\_120725164543540 FN597028\_120725164606045  
FN597030\_120725164625300 FN597032\_120725164722623 FN597034\_120725164752719 FN597036\_120725164812988  
FN597038\_120725164832300 FN597039\_120725164904489 FN597040\_120725164929125 FN597042\_120725164953101  
FN597044\_120725165018221 FN597046\_120725165040791 FN597015\_120725164150286

**Download modified queries**

- [SRR031112.fastq \(583.0 MB\)](#)

**Download merged pileup file**

- [merged.var.flt.vcf \(2.5 GB\)](#)
- [merged.sam \(2.0 GB\)](#)

**Download wgs file**

- [out\\_WGS.fasta \(95.7 MB\)](#)

コンティグファイルのダウンロード

**Position errors** **Map ratio** **Depth, Coverage**

**Time**

Wait time	Start time	End time
0: 1:38	2012-08-01 11:04:45	2012-08-01 11:54:30

# コンテイグファイル

---

```
>ENTRY_1_1
NNNTAATTGGAGTTAATTGCCTAATGAAAAAA
//  
>ENTRY_1_2
AAGAACAGAACAGTGTAACTATGAAACTAAGAT
//  
>ENTRY_1_3
AAGCCCTAACTAACACCTAAACACATTAATTAA
//  
>ENTRY_1_4
AAGCTTAGAGTCCTTTAATTTTAAAAACAAAATC
//  
>ENTRY_1_5
GCATGAGTATAACCCATCTCTCTATCATTGACAG
//  
>ENTRY_1_6
ATGGATCATTCTACAAAATACTCTTACAATACCG
//  
>ENTRY_1_7
GTNCNGCCTAGGNGNGNNNNACCCATGTACAAAAAGGGAAAANAAAAATTGAAAAA
//  
>ENTRY_1_8
TNNTTAGGTTGTTCTCTGNANCNNNNNNNNNNN
//  
>ENTRY_1_9
TCCNNNGATCTGNNACNNNNNNNNNNCTNNNNNNNNN
//  
>ENTRY_1_10
GTCNCCAAAATGTGGGCAGTTATTTGGCTGCTTATA
//  
>ENTRY_1_11
TGATTCTCACTGTATCACCTAATGTCTTATGTATGG
//  
>ENTRY_1_12
GAGTTGTTTCAAAGGAGGTTCAGTGCTTGTGG
//  
>ENTRY_1_13
TATAAAGTTTAGCTCAAGTCTGGCTTAGAGACTC
//  
>ENTRY_1_14
TATAAAGTTTAGCTCAAGTCTGGCTTAGAGACTCATTNTGGTAAAAATGGATTAGGT
AGCATTAGGG
//
```

# マッピング結果のビューア

## ◆ Tablet

○ <http://ics.hutton.ac.uk/tablet/>

○ windows, Mac OSX, Linux

## ◆ IGV

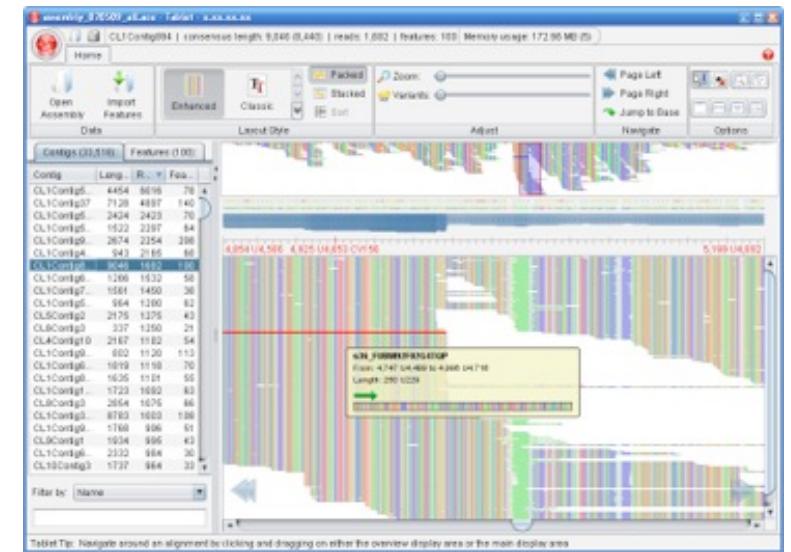
○ <http://www.broadinstitute.org/software/igv/home>

○ windows, Mac OSX, Linux

## ◆ samtools tview

○ <http://samtools.sourceforge.net>

○ Mac OSX, Linux



# p-galaxyへのリンク

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 : DRA000001

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

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
<input type="checkbox"/>	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.)

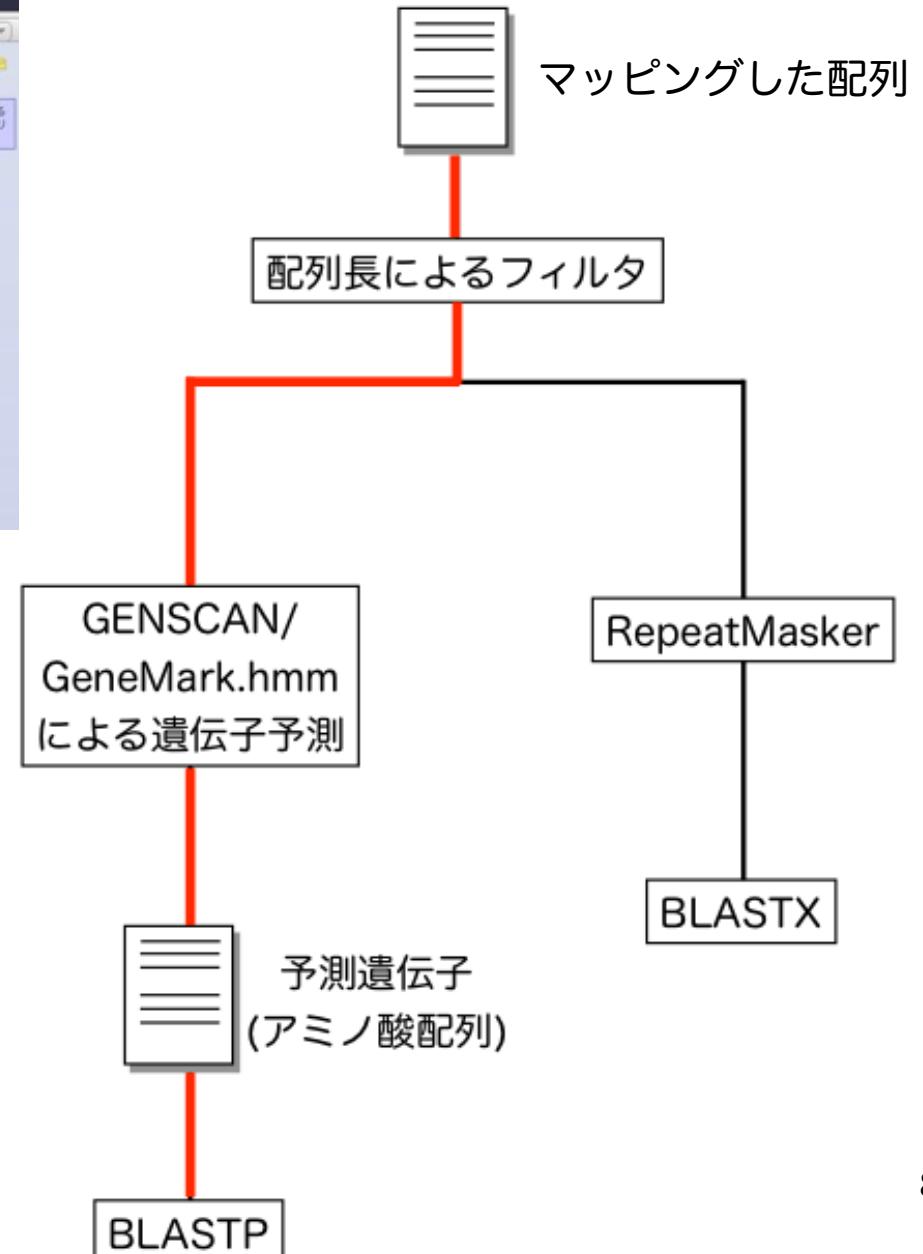
DELETE NEXT

# p-Galaxyを使った配列アノテーション

The screenshot shows the p-Galaxy web interface. At the top, there's a navigation bar with 'Galaxy / DDBJ' and links for 'Analyze Data', 'Workflow', 'Shared Data', 'P-galaxy Manual', 'Help', and 'User'. On the left, a sidebar lists various tools and analysis types. A central panel displays a workflow titled 'WWFSMD?' with the sub-instruction 'grow noodly appendages...'. Below the workflow, it says 'usegalaxy.org'. A message at the bottom states: 'This project is supported in part by NSF, NHGRI, and the Huck Institutes of the Life Sciences.'

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

(開発中のサービスのため、一部サービスが  
使えなかったり、一時的にアクセスできなくな  
ったりする可能性があります)



# Galaxyとは？

- ◆ ゲノムなど生物学データを対象とした、データ解析ワークフローの共有・公開のためのプラットフォーム
- ◆ オリジナルはペンシルバニア州立大学、エモリー大学を中心としたGalaxy teamが開発
- ◆ 独自のツールを追加して公開可能

## ● DBCLS Galaxy

- テキスト系ツール

## ● p-Galaxy

- 次世代解析ツール



# ダウンロードしたファイルをアップロード

The screenshot shows the Galaxy / P-GALaxy web application. The top navigation bar includes links for Analyze Data, ワークフロー, Shared Data, Visualization, P-galaxy Manual, Help, User, and a grid icon indicating 0 bytes. The left sidebar contains a search tools section and a list of data sources under Work Flow, including Get Data (with 'Upload File from your computer' highlighted), UCSC Main table browser, UCSC Test table browser, UCSC Archaea table browser, BX table browser, EBI SRA ENA SRA, Get Microbial Data, BioMart Central server, BioMart Test server, CBI Rice Mart rice mart, GrameneMart Central server, modENCODE fly server, Flymine server, Flymine test server, modENCODE modMine server, MouseMine server, Ratmine server, YeastMine server, and metabolicMine server. The main content area is titled 'Upload File (version 1.1.3)'. It features sections for 'File Format' (Auto-detect), 'File' (containing 'out\_WGS\_pino\_bwa.fasta'), 'URL/Text' (empty), 'Files uploaded via FTP' (instructions to create or log in to a Galaxy account), 'Convert spaces to tabs' (checkbox 'Yes' checked), 'Genome' (dropdown menu set to 'unspecified (?)'), and a 'Execute' button. A status bar on the right indicates 'ヒストリー' (History) with 'Unnamed history' and '0 bytes'. A note in the history panel says 'ヒストリーは空です。解析をはじめるには、左パネルの'データ取得'をクリック'.

ツールを選択

メイン操作画面

解析データリスト 86

# データの確認

The screenshot shows the Galaxy / P-GALaxy web application. On the left, there's a sidebar with various 'Get Data' options like 'Upload File from your computer', 'UCSC Main table browser', etc. The main area displays a large dataset with entries labeled ENTRY\_1\_1 through ENTRY\_1\_14. A yellow warning box at the top states: 'This dataset is large and only the first megabyte is shown below.' with links to 'Show all' or 'Save'. To the right is a 'History' panel titled 'Unnamed history' containing a single item: '1: out WGS pino bwa.fasta'. A red arrow points from the text '目アイコンをクリックすると、真ん中にデータが表示される' to the eye icon next to the history item. Another red arrow points from the text '操作が成功すると緑色 操作が失敗すると赤色' to the green success status bar.

This dataset is large and only the first megabyte is shown below.  
[Show all](#) | [Save](#)

```
>ENTRY_1_1
NNNTAAATTGGAGTTAATTGCCTAATGAAAAAN
//
>ENTRY_1_2
AAGAACAGAANCAAGTGCTAACTATGAAACTAAGAT
//
>ENTRY_1_3
AAGCCCTAACTAACACCTAACACATTAATTAAATT
//
>ENTRY_1_4
AAGCTTAGAGTCTTTAATTTTAAAAACAAAATC
//
>ENTRY_1_5
GCATGAGTATAACCCATCTCTCTATCATTGACAG
//
>ENTRY_1_6
ATGGATCATTCTACAAAATCTCTTACAATACCG
//
>ENTRY_1_7
GTNCNGCCTTAGGNGNGNNNNACCATGTACAAAAAGGGAAAANAAAAAATTGAAAAA
//
>ENTRY_1_8
TNTTTAGGTTGTTCTCTGNANCNNGNNNNNNNNNN
//
>ENTRY_1_9
TCCNNNGATCTGNNAACNNNNNNNNCTNNNNNNNNNN
//
>ENTRY_1_10
GTCNCCAAAATGTGGGCAGTTATTGGCTGCTTATA
//
>ENTRY_1_11
TGATTCTCACTGTATCACCTAATGTCTTATGTATGG
//
>ENTRY_1_12
GAGTTGTTTCAAAGGAGGTTTCAGTGCTTGTGCGG
//
>ENTRY_1_13
TATAAAGTTTAGCTCAAGTCTGGCTTAGAGACTC
//
>ENTRY_1_14
TATAAAGTTTAGCTCAAGTCTGGCTTAGAGACTCATTNTGGTAAAAATGGATTAGGT
```

ヒストリー

Unnamed history  
91.3 MB

1: out WGS pino bwa.fasta

操作が成功すると緑色  
操作が失敗すると赤色

# コンティグの配列長でフィルタリング

ツール Work Flow を開いて

ANNOTATION FOR DE NOVO ASSEMBLED SEQ.

**FASTA File Length Filter**

Gene Prediction Choose GENSCAN or GeneMark.hmm

Gene Prediction to FASTA Converts GENSCAN or GeneMark.hmm output file to FASTA format

transcriptsToOrfs Trinity Transcripts to Candidate Peptides

RepeatMasker

BLASTP

FASTA File Length Filter (version 1.0.0)

Input FASTA File:  
1: out\_WGS\_pino\_bwa.fasta

Base length of data removing from input file:  
500

The sequence data that length is the same as or less than this value is removed.

Execute

ヒストリー

Unnamed history  
91.3 MB

1: out\_WGS\_pino\_bwa.fasta



ツール

ANNOTATION FOR DE NOVO ASSEMBLED SEQ.

FASTA File Length Filter

Gene Prediction Choose GENSCAN or GeneMark.hmm

Gene Prediction to FASTA Converts GENSCAN or GeneMark.hmm output file to FASTA format

transcriptsToOrfs Trinity Transcripts to Candidate Peptides

RepeatMasker

This dataset is large and only the first megabyte is shown below.  
[Show all](#) | [Save](#)

```
>ENTRY_1_986
GAAAAGTTAACATGACGGCTATGAAAATACATATGCTCTACACAAACGGACGTAAGAAGATCCTCGTCCAATG
>ENTRY_1_2288
TGTTTGCAATTGTAGGGAGTCAGTTAGGAGTCGGTTGAGTCTGGAGTCACATTCTGGARGAGAGTTCGAGGACGATTG
>ENTRY_1_2292
GGATTGGTTGTCCAAGATCCATTGAGACTATACAGCACAGTCATGAGAGCACATGGGATAGATGATGRCAGTTGG
>ENTRY_1_2910
AAGATATCCGAGAAATGGATATCATATTGAAACTATGAATGAAGATAATGTAGAATATCTTATATTANTCCATT
```

ヒストリー

Unnamed history  
93.3 MB

2: FASTA File Length Filter on data 1

1: out\_WGS\_pino\_bwa.fasta

# ORF予測

Galaxy / P-GALAXY Analyze Data ワークフロー Shared Data Visualization P-galaxy Manual Help User Using 93.3 MB

ツール

ANNOTATION FOR DE NOVO ASSEMBLED SEQ.

FASTA File Length Filter

Gene Prediction Choose GENSCAN or GeneMark.hmm

Gene Prediction to FASTA Converts GENSCAN or GeneMark.hmm output file to FASTA format

transcriptsToOrfs Trinity Transcripts to Candidate Peptides

RepeatMasker

Gene Prediction (version 1.0.0)

Select a gene prediction tool: GENSCAN

Parameter file: Zea mays

For appropriate organism

Sequence file: 2: FASTA File Length Filter on data 1

FastA formatted file

Execute

ヒストリー

Unnamed history 93.3 MB

2: FASTA File Length Filter on data 1

1: out WGS pino bwa.fasta

Galaxy / P-GALAXY Analyze Data ワークフロー Shared Data Visualization P-galaxy Manual Help User Using 94.6 MB

ツール

ANNOTATION FOR DE NOVO ASSEMBLED SEQ.

FASTA File Length Filter

Gene Prediction Choose GENSCAN or GeneMark.hmm

Gene Prediction to FASTA Converts GENSCAN or GeneMark.hmm output file to FASTA format

transcriptsToOrfs Trinity Transcripts to Candidate Peptides

RepeatMasker

BLASTP

BLASTX

PHYLOGENETIC ANALYSIS

sam to fasta for get mapping fasta

```
>ENTRY_1_7974|GENSCAN_predicted_peptide_1155_aa
MSEMRRSSEAETAENQRTGPYQCEEVFHVTDSNDRECGQVYCNRCNEDVGLYLX

>ENTRY_1_7974|GENSCAN_predicted_CDS_11165_bp
atgtcttagatcgccggaggactcgaggagtcggatcgacagcggagaatcgaggactgggcca
tatgggcagtgcgaggaggcttccatgtactatcgatagcaacgatcgcgagtgtggccaa
gtttatttcaataggtaatggacgttgccttacctcggn
GENSCAN 1.0 Date run: 1-Jul-114 Time: 21:21:08

Sequence ENTRY_1_7981 : 538 bp : 47.76% C+G : Isochore 1 ( 0 - 100 C+G%)

Parameter matrix: Maize.smat

Predicted genes/exons:
```

Gn.Ex Type S.Begin ...End .Len Fr Ph I/Ac Do/T CodRg P.... Tscr..

-----

1.01 Init- 135 78 58 0 1 86 101 29 0.725 8.06

Predicted peptide sequence(s):

# 予測結果からペプチドの配列を抽出

Galaxy / P-GALAXY Analyze Data ワークフロー Shared Data Visualization P-galaxy Manual Help User Using 94.6 MB

ツール

ANNOTATION FOR DE NOVO ASSEMBLED SEQ.

FASTA File Length Filter

Gene Prediction Choose GENSCAN or GeneMark.hmm

**Gene Prediction to FASTA Converts GENSCAN or GeneMark.hmm output file to FASTA format**

transcriptsToOrfs Trinity Transcripts to Candidate Peptides

Gene Prediction to FASTA (version 1.0.0)

Input Gene Prediction File: 3: Gene Prediction on data 2

Select extracting sequences: Peptide sequences

Execute

ヒストリー

Unnamed history 94.6 MB

3: Gene Prediction on data 2 / X

2: FASTA File Length Filter on data 1 / X

1: out WGS pino bwa.fasta / X

Galaxy / P-GALAXY Analyze Data ワークフロー Shared Data Visualization P-galaxy Manual Help User Using 94.7 MB

ツール

ANNOTATION FOR DE NOVO ASSEMBLED SEQ.

FASTA File Length Filter

Gene Prediction Choose GENSCAN or GeneMark.hmm

Gene Prediction to FASTA Converts GENSCAN or GeneMark.hmm output file to FASTA format

transcriptsToOrfs Trinity Transcripts to Candidate Peptides

RepeatMasker

BLASTP

BLASTX

PHYLOGENETIC ANALYSIS

```
>ENTRY_1_3485IGENSCAN_predicted_peptide_1I6_aa  
MNLLLL  
>ENTRY_1_3486IGENSCAN_predicted_peptide_1I6_aa  
MNLLLL  
>ENTRY_1_3487IGENSCAN_predicted_peptide_1I6_aa  
MNLLLL  
>ENTRY_1_7974IGENSCAN_predicted_peptide_1I55_aa  
MSEMRRSSEAETAENQRTGPYGQCEEVFHVTDSNDRECGQVYCNRCNEDVGLYLYX  
>ENTRY_1_7981IGENSCAN_predicted_peptide_1I20_aa  
MGMGKTSETGRDSEVVEKNX  
>ENTRY_1_8473IGENSCAN_predicted_peptide_1I2_aa  
MX  
>ENTRY_1_8699IGENSCAN_predicted_peptide_1I15_aa  
MAITLRAADRDRGX  
>ENTRY_1_11502IGENSCAN_predicted_peptide_2I67_aa  
XRKGGSASGEPGGGSASPESGSRALTRRRVRCSSRPEKRAWAARGWVSGSGALGQIGDLLQALLAV  
>ENTRY_1_15778IGENSCAN_predicted_peptide_1I81_aa  
MLVGGRRGWRRRRVAGKSLTRPHAHGAPGAFTKVGDLLQALLLVWSVSEKDVGNSPENFAGGECFLTTIRLTGKYWEMEK
```

ヒストリー

Unnamed history 94.7 MB

4: Gene Prediction to FASTA / X on data 3

3: Gene Prediction on data 2 / X

2: FASTA File Length Filter / X on data 1

1: out WGS pino bwa.fasta / X

# 予測ORF配列の機能アノテーション

Galaxy / P-GALAXY Analyze Data ワークフロー Shared Data Visualization P-galaxy Manual Help User Using 94.7 MB

ツール

ANNOTATION FOR DE NOVO ASSEMBLED SEQ.

FASTA File Length Filter

Gene Prediction Choose GENSCAN or GeneMark.hmm

Gene Prediction to FASTA Converts GENSCAN or GeneMark.hmm output file to FASTA format

transcriptsToOrfs Trinity Transcripts to Candidate Peptides

RepeatMasker

**BLASTP**

BLASTX

BLASTP (version 1.0.0)

Query file: 4: Gene Prediction to FASTA on data 3

Select database: nr

Expectation value: 0.00001  
ex.) 0.00001 or -20 (as e-20)

Execute

ヒストリー

Unnamed history 94.7 MB

4: Gene Prediction to FASTA on data 3

3: Gene Prediction on data 2

2: FASTA File Length Filter on data 1

1: out WGS pino bwa.fasta

Galaxy / P-GALAXY Analyze Data ワークフロー Shared Data Visualization P-galaxy Manual Help User Using 95.6 MB

ツール

ANNOTATION FOR DE NOVO ASSEMBLED SEQ.

FASTA File Length Filter

Gene Prediction Choose GENSCAN or GeneMark.hmm

Gene Prediction to FASTA Converts GENSCAN or GeneMark.hmm output file to FASTA format

transcriptsToOrfs Trinity Transcripts to Candidate Peptides

RepeatMasker

**BLASTP**

Sequences producing significant alignments: (bits) Value

spIP64956IY2253\_MYCBO Uncharacterized protein Mb2253c OS=Mycobac... 33 0.069  
spIP64955IY2228\_MYCTU Uncharacterized protein Rv2228c/MT2287 OS=... 33 0.069

>spIP64956IY2253\_MYCBO Uncharacterized protein Mb2253c  
OS=Mycobacterium bovis (strain ATCC BAA-935 / AF2122/97) GN=Mb2253c PE=4 SV=1  
Length = 364

Score = 33.5 bits (75), Expect = 0.069, Method: Compositional matrix adjust.  
Identities = 15/35 (42%), Positives = 24/35 (68%)

Query: 17 ASNNEAEYEAILSGLDLALALSVSKLRIYSDSLQV 51  
A+NN AEY +++GLD A+ L ++ + DS+LV  
Sbjct: 43 ATNNVAEYRGLIAGLDDAVKLGATEAAVLMDSKLV 77

ヒストリー

Unnamed history 95.6 MB

16: BLASTP error/warning reports

15: BLASTP on data 4

4: Gene Prediction to FASTA on data 3

3: Gene Prediction on data 2

2: FASTA File Length Filter on data 1 91

1: out WGS pino bwa.fasta



©2012 統合データベース講習会 Licensed Under CC 表示 2.1

# DDBJ Pipeline 参考資料

## ◆ DDBJing講習会

<http://www.ddbj.nig.ac.jp/ddbjing/dl.html>

### ◆ 「第29回 DDBJing 講習会 in 三島」 2014年6月12日

- » テーマ : DDBJ へのデータ登録
- » 場 所 : 国立遺伝学研究所 生命情報・DDBJ 研究センター4階
- » 主 催 : 日本DNAデータバンク (DDBJ)

タイトル	講師	資料
DDBJ, スパコンの紹介, 大量配列情報解析	中村 保一 (国立遺伝学研究所)	  
公共データベースを使い倒した知のめぐりのよい生命科学研究	坊農 秀雅 (ライフサイエンス統合データベースセンター)	  
NIG スーパーコンピュータの紹介と見学	石川 直史 (国立遺伝学研究所)	  
NBDCヒトデータベースを介したJapanese Genotype-phenotype Archive のデータ共有の審査過程と登録手続き	箕輪 真理 (JST バイオサイエンスデータベースセンター)	  
BioProject と BioSample の紹介	児玉 悠一 (DDBJ)	  
DDBJ Sequence Read Archive の紹介	古屋 典子 (DDBJ)	  
Mass Submission System の紹介	李 慶範 (DDBJ)	  
DDBJ Nucleotide Sequence Submission System の紹介	小菅 武英 (DDBJ)	  

### ◆ 「第28回 DDBJing 講習会 in 新潟」 2013年11月19日

- » テーマ : NGS解析Galaxy特集とDDBJへの登録
- » 場 所 : 新潟大学 駅南キャンパス「ときめいと」 (JR 新潟駅南口より徒歩3分) [アクセス](#)
- » 主 催 : 新潟大学 (工学部情報工学科 阿部貴志), 日本DNAデータバンク (DDBJ)

タイトル	講師	資料	USTREAM
DDBJ, スパコンの紹介, 大量配列情報解析	中村 保一 (国立遺伝学研究所)	 	
Galaxy入門とワークフロー構築	山口 敦子 (ライフサイエンス統合データベースセンター)	 	
数百人規模オミックス解析のための Galaxy 活用術	八谷 剛史 (岩手医科大学 いわて東北メディカル・メガバンク機構)	準備中	
新世代シークエンスデータの登録	児玉 悠一 (DDBJ)	 	
DDBJ Pipeline の紹介と実習	長崎 英樹 (国立遺伝学研究所)	 	
大量登録システム(MSS) の紹介 (第27回DDBJing 講習会 資料)			

# MiGAP

---

- ◆ 微生物ゲノムのアノテーションワークフロー
- ◆ MiGAPで付加するアノテーション情報
  - ORF (CDS)の同定
    - de novo 予測 ⇒ MetaGeneAnnotator, Glimmer, Augustus
  - rRNAの予測
    - de novo 予測 ⇒ RNAmmer
    - 既知rRNA配列との配列類似性
  - tRNAの予測
    - de novo 予測 ⇒ tRNAScan-SE
  - ORFの機能予測
    - BLAST

# MiGAPウェブサイト

<http://www.migap.org/>

The screenshot shows the MiGAP website homepage. On the left sidebar, there are links for 'トップメニュー' (Top Menu) including 'MIGAPについて', 'ヘルプ', 'お知らせ', 'バイオラインについて', 'MIGAPサーバの運用主体', 'MIGAP引用・関連文献リスト', and '謝辞'. Below that is a 'ログイン' (Login) section for '旧バイオライン' (Old BioLine) with fields for 'ユーザ名' (Username) and 'パスワード' (Password), and a 'ログイン' (Login) button. At the bottom of this sidebar is a large blue button labeled 'アカウントを取得する' (Get Account). On the right side, there are two main news sections: 'MiGAPサービス再開しました' (MiGAP service has been restored) dated June 26, 2014, and '国立遺伝学研究所空調電源工事に伴うMiGAP投入受付停止のお知らせ' (Notice of MiGAP submission processing stop due to air conditioning power supply work at the National Institute of Genetics) dated June 19, 2014. Both news items include a small British flag icon. To the right of these news sections is a '最新ニュース' (Latest News) sidebar with links to various news items about MiGAP's status and its impact on bioinformatics research.

■ トップメニュー

- MIGAPについて
- ヘルプ
- お知らせ
- バイオラインについて
- MIGAPサーバの運用主体
- MIGAP引用・関連文献リスト
- 謝辞

■ バイオラインにログイン

■ 旧バイオラインにログイン  
公開中です

■ フォーラムにログイン

ユーザ名

パスワード

オートログイン

**ログイン**

パスワードを忘めましたか?  
ユーザ名を忘めましたか?

■ アカウントを取得する

**アカウントを取得する**

Home

### MiGAPサービス再開しました

2014年6月26日(木曜日)09:58 | 最終更新 2014年6月26日(木曜日)10:00 | 作者: 齊藤 仁浩 |

MiGAPサービスを再開しました。空調電源工事のため停止していたMiGAPサービスへのジョブ投入を再開しました。

### 国立遺伝学研究所空調電源工事に伴うMiGAP投入受付停止のお知らせ

2014年6月19日(木曜日)12:59 | 最終更新 2014年6月19日(木曜日)13:43 | 作者: 齊藤 仁浩 |

国立遺伝学研究所所内空調電源工事のため、MiGAPへの投入受付を以下の期間停止いたします。

**ジョブ投入受付停止時刻： 2014年6月23日（月）12:00**

なお、実行待ちジョブが多数ある場合は、ジョブ投入受付停止時刻を早める場合があります。

また、ジョブ投入受付停止時刻以前に投入されたジョブであっても、その処理が以下の時刻までに実行終了しない場合は、それらジョブは強制的にキャンセルされます。

**残存ジョブ強制終了時刻： 2014年6月23日（月）17:00**

サービス再開予定は以下の通りです。

**ジョブ投入受付再開予定時刻： 2014年6月26日（木）10:00**

### MiGAPサービス再開しました

2014年5月21日(水曜日)13:20 | 最終更新 2014年5月21日(水曜日)13:22 | 作者: Administrator |

2014年5月21日13:20 MiGAPサービスを再開しました。停電のため停止していたMiGAPサービスへのジョブ投入を再開しました。

■ 最新ニュース

- MiGAPサービス再開しました
- 国立遺伝学研究所空調電源工事に伴うMiGAP投入受付停止のお知らせ
- MiGAPサービス再開しました
- 送電事故による停電によりMiGAPは停止しております
- MiGAPサービス再開しました
- 停電のため、現在MiGAPは停止しています
- MiGAPサービス再開しました
- 遺伝研所内停電に伴うMiGAP投入受付停止のお知らせ
- 2013年11月後半に投入されたジョブへの解析結果に不具合がありました（お詫び）
- 遺伝研所内停電に伴うMiGAP投入受付停止のお知らせ

■ 閲覧ランキング

- MiGAPとは？
- MiGAP講習会のお知らせ
- フォーラムヘルプ
- MiGAPサーバの運用主体
- 遺伝研所内電源工事に伴うMiGAP投入受付停止のお知らせ
- b-MiGAPチュートリアル
- MiGAPを担当しているのは？
- 謝辞
- Introduction & Practice (as of May 2012)
- 投入塩基長によるバイオラインへの投入間隔が変更されました

検索...

# ユーザレベルの設定

The screenshot shows the MiGAP interface with a blue header bar containing 'Logout', 'Help', and 'Contact Us'. Below the header is a navigation menu with 'Pipe Line' (selected), 'History', 'Change User Level' (highlighted with a red box), 'Current Process', and 'Set'. To the right is a 'Change User Level' section with three radio button options: 'b-MiGAP' (selected), 's-MiGAP', and 'g-MiGAP'.

## ◆ b-MiGAP

- ツール・パラメータ固定（ブロンズ）

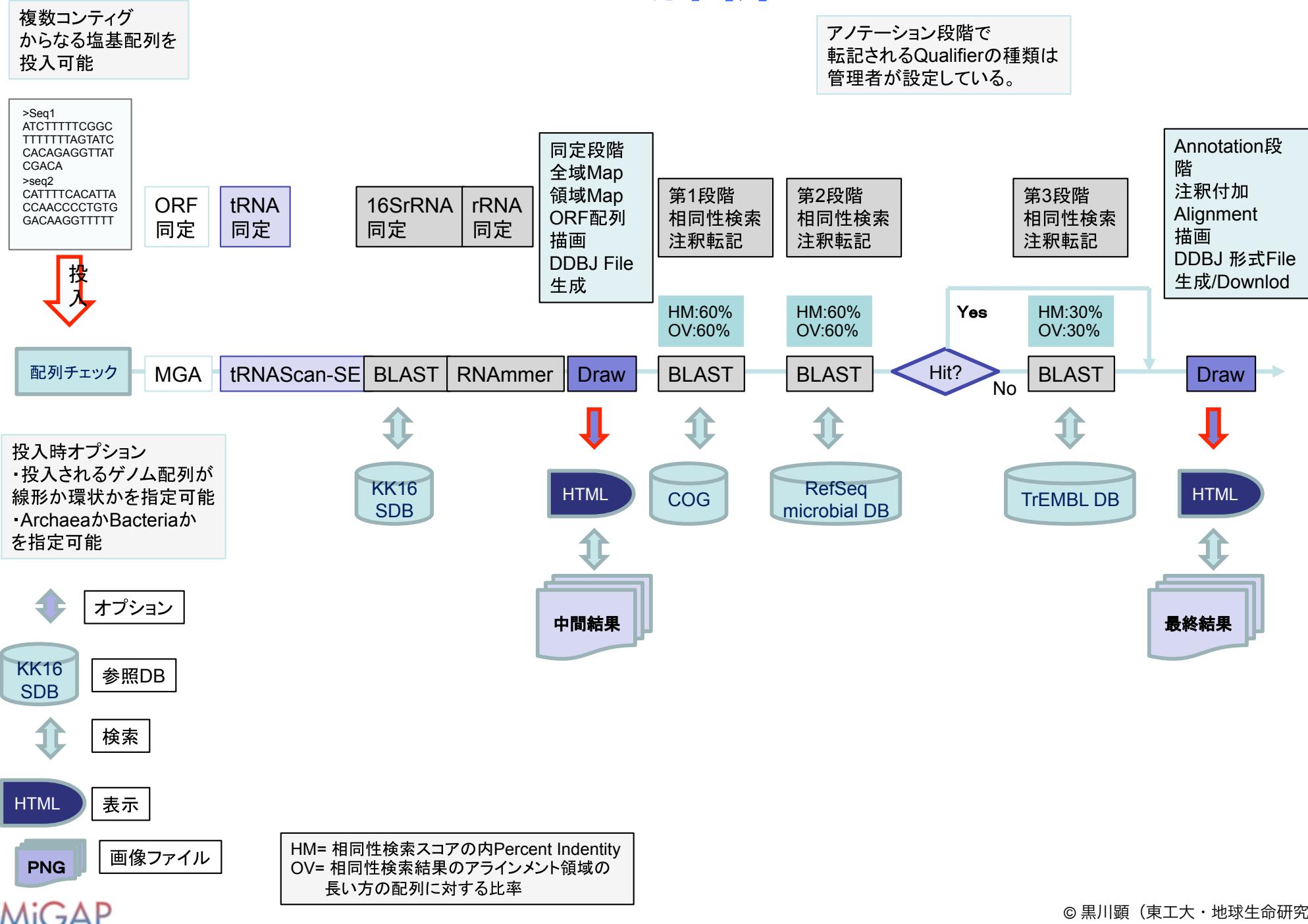
## ◆ s-MiGAP

- パラメータ設定が可能（シルバー）

## ◆ g-MiGAP

- DBやツール、パラメータ設定が可能（ゴールド）

# b-MiGAP 解析フロー



# MiGAPデータ入力

**MiGAP Microbial Genome Annotation Pipeline ver2.18**

Logout Help Contact Us LDAP\_kawano (b-MiGAP) 2014/07/01 20:45:06 [View Menu] [Hide Menu]

Pipe Line  
Pipe Line History  
Change  
User Level  
Current Process

Input Pipe Line [Running:3 Waiting:0]

Pipe Line Name:

Upload Filename:  ファイルを選択 選択されていません ← アセンブル済みファイルをアップロード  
or paste data in box below.([Sample data](#)) ← ボックスに入力

Linear Circular Bacteria Archaea Eukarya \*FUNGI\* transl\_table: 11

Run Clear ← リニア or 環状 ← 生物ドメインの選択

# MiGAP結果

**MiGAP Microbial Genome Annotation Pipeline ver2.18**

Logout Help Contact Us LDAP\_kawano (b-MiGAP) 2014/07/01 22:14:58 [View Menu] [Hide Menu]

Pipe Line History

Pipe Line History List LDAP\_kawano | Hidden Data List Contig [1/1] contig 2014/07/01 20:55:27 kousyuukai pre running hokudai multi contig(whole) hoge 2013/10/15 15:34:34 2013/10/15 14:57:52 2013/10/15 14:56:49 2013/10/15 14:55:48 2013/10/15 14:39:41 2013/10/15 14:37:31 2013/10/15 14:36:28 2013/10/15 14:35:25 2012/08/02 16:52:25(whole) 2012/08/01 14:38:37 Solanum lycopersicum DNA, contig: SISBM\_S00131\_04

Basic Information

Filename: direct  
Contig: 1  
Total Length: 10530  
rRNA: 3  
tRNA: 2  
CDS: 3  
RBS: 3  
Run: 2014/07/01 20:55:27

ORF & RNA Extract

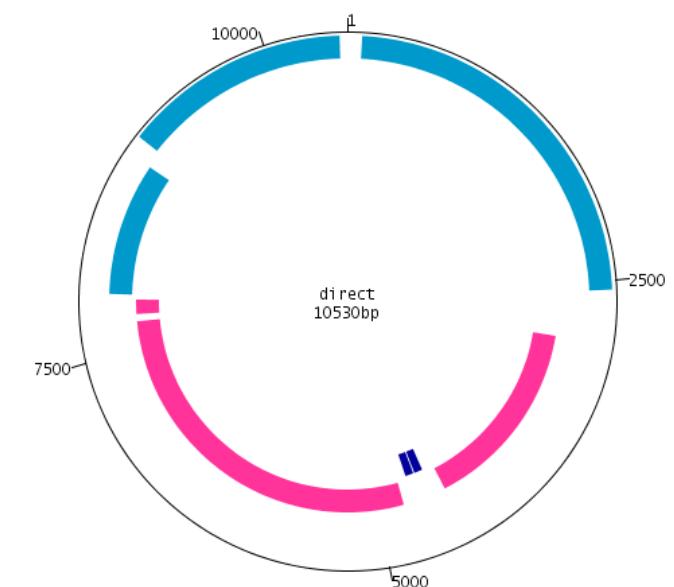
Start: 2014/07/01 20:55:29  
End: 2014/07/01 20:56:24  
Software: MetaGeneAnnotator 1.0  
tRNAscan-SE 1.23  
NCBI BLAST 2.2.18  
RNAmmer 1.2

Annotation

Start: 2014/07/01 20:56:24  
End: 2014/07/01 21:03:10  
Software: NCBI BLAST 2.2.18  
DB: COG[20030417]:3  
RefSeq release56[20130314]:3  
TrEMBL release2013\_03[20130306]:0

Download

Log File [pipeline.log](#)  
N.A.: [result-na.fasta.tar.gz](#)  
A.A.: [result-aa.fasta.tar.gz](#)



# MiGAP参考資料

---

- 統合TV

- ✓ 微生物ゲノムアノテーションツールMiGAP

- ▶ 開発者である黒川先生による講習会の動画
    - <http://youtu.be/ujxl6LJlbUE>
    - <http://togotv.dbcls.jp/20131024.html>

- ✓ MiGAPの使い方～導入と基本操作

- ▶ 動画による解説
    - <http://youtu.be/oXAEZgoc5Eo>
    - <http://togotv.dbcls.jp/20100624.html>

# 参考サイト

---

## ◆ NGS Surfer's wiki



- <https://cell-innovation.nig.ac.jp/wiki/tiki-index.php>

## ◆ Q and A

- ライフサイエンスQA (日本語)



- <http://qa.lifesciencedb.jp/>

- BioStar (英語)



- <https://www.biostars.org/>

- SEQanswers (英語)



- <http://seqanswers.com/>

## • (Rで)塩基配列解析

- [http://www.iu.a.u-tokyo.ac.jp/~kadota/r\\_seq.html](http://www.iu.a.u-tokyo.ac.jp/~kadota/r_seq.html)