

INSDC の塩基配列データベースを知る・使う・その舞台裏

児玉 悠一

国立遺伝学研究所 バイオデータ研究拠点 (BSI)

自己紹介



国立遺伝学研究所 バイオデータ研究拠点 (BSI) コーディネータ

1. INSDC 塩基配列データベースの利用方法 - 検索と登録

13:40～14:50 (70分)

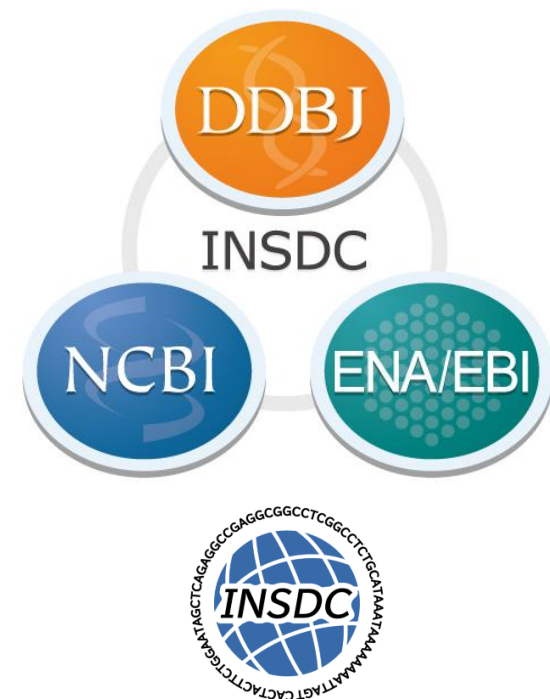
2. 国際塩基配列データベースの舞台裏 - 増大するデータ・透明化・社会的責任

15:00～15:50 (50分)

1. INSDC 塩基配列データベースの利用方法 - 検索と登録

INSDC: 国際塩基配列データベース協力体制

1980	EMBL Data Library が EMBL ハイデルベルグ拠点で開始
1982	GenBank がロスアラモス研究所で公開
1987	DDBJ が国立遺伝学研究所で正式に活動開始
1990年代	GenBank が NCBI/NIH に移管 EMBL-Bank (Data Library) が ENA/EMBL-EBI (英国) に移管
2005	国際塩基配列データベース協力体制を INSDC と総称 次世代シーケンサの登場
2009	SRA
2010年頃	BioProject
2012年頃	BioSample
2027	DDBJ 40周年



データを見て辿って、それぞれのデータベースを理解する

► J Adv Res. 2024 Oct 28;75:863–874. doi: [10.1016/j.jare.2024.10.023](https://doi.org/10.1016/j.jare.2024.10.023) 

Chromosome-scale assembly with improved annotation provides insights into breed-wide genomic structure and diversity in domestic cats

[Yuki Matsumoto](#)^{a,b,1}, [Claire Yik-Lok Chung](#)^{c,1}, [Sachiko Isobe](#)^{d,1}, [Mika Sakamoto](#)^{e,1}, [Xiao Lin](#)^c, [Ting-Fung Chan](#)^c, [Hideki Hirakawa](#)^d, [Genki Ishihara](#)^a, [Hon-Ming Lam](#)^c, [Shinobu Nakayama](#)^d, [Shigemi Sasamoto](#)^d, [Yasuhiro Tanizawa](#)^e, [Akiko Watanabe](#)^d, [Kei Watanabe](#)^a, [Masaru Yagura](#)^e, [Yoshihito Niimura](#)^f, [Yasukazu Nakamura](#)^{e,*}

► Author information ► Article notes ► Copyright and License information

PMCID: PMC12789764 PMID: [39490737](https://pubmed.ncbi.nlm.nih.gov/39490737/)

Illumina and PacBio sequencing

Genomic DNA was extracted from Senzu's blood using a DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany). Sequence libraries were prepared using SMRTbell Express Template Prep Kit 2.0 (PacBio, Menlo Park, CA, USA). Library size was selected using BluePippin (Sage Science, Beverly, MA, USA) to remove DNA fragments smaller than 30 kb. The library was sequenced on a Sequel system (PacBio) using 28 SMRT cells (**PRJDB9879**). The sequence reads were assembled using FALCON v.1.8.8 (c) with a length cutoff of 34 kb to generate primary contig sequences and associated contigs representing alternative alleles. Haplotype-resolved assemblies, that is, primary contigs and haplotigs, were generated using FALCON Unzip ver. 1.8.8 [23]. Potential errors in the resulting primary contig sequences were corrected twice using ARROW ver. 2.2.1, implemented on the SMRT Link (v 5.0; PacBio). Paired-end libraries were constructed for Senzu's DNA and sequenced using an Illumina HiSeqX system (Illumina), with a read length of 151 bp (**PRJDB9879**). The obtained short reads were used for further error correction of contig sequences using Pilon [24]. Short-read data were also used for genome size estimation using Jellyfish ver. 2.1.4 [25].



Name: Senzu

NCBI で検索

NIH

National Library of Medicine

National Center for Biotechnology Information

Log in

All Databases

PRJDB9879

Search

NCBI Home

Resource List (A-Z)

All Resources

Chemicals & Bioassays

Data & Software

DNA & RNA

Domains & Structures

Genes & Expression

Genetics & Medicine

Genomes & Maps

Homology

Literature

Proteins

Sequence Analysis

Taxonomy

Training & Tutorials

Variation

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News & Blog](#) | [Visit the Beta Homepage](#)

Submit

Deposit data or manuscripts into NCBI databases

Download

Transfer NCBI data to your computer

Learn

Find help documents, attend a class or watch a tutorial

Develop

Use NCBI APIs and code libraries to build applications

Analyze

Identify an NCBI tool for your data analysis task

Research

Explore NCBI research and collaborative projects

Popular Resources

PubMed

Bookshelf

PubMed Central

BLAST

Nucleotide

Genome

SNP

Gene

Protein

PubChem

NCBI News

Upcoming Transparency Compliance M of the 2024 N

As shared in s

Now Available

RefSeq releas available onlin You can acce

Search NCBI

PRJDB9879

Search

Results found in 6 databases

BIOPROJECT

Genome sequencing project of American Shorthair cat

The domestic cat (Felis catus) is one of the most popular companion animals in the world. Modern cat breeds have variations in traits such as coat color, tail length and ear shape as well as ...

PRJDB9879

Genomes

Literature

Bookshelf0

MeSH0

NLM Catalog0

PubMed0

PubMed Central1

Genes

Gene0

GEO DataSets0

GEO Profiles0

Proteins

Conserved Domains0

Identical Protein Groups0

Protein0

Protein Family Models0

Structure0

Genomes

Assembly / Genome NCBI Datasets1

BioCollections0

Clinical

ClinicalTrials.gov0

ClinVar0

PubChem

BioAssays0

Compounds0

<https://ncbi.nlm.nih.gov/>

BioProject で俯瞰

Felis catus strain:American Shorthair (domestic cat) Accession: PRJDB9879 ID: 63448

Genome sequencing project of American Shorthair cat

The domestic cat (*Felis catus*) is one of the most popular companion animals in the world. Modern cat breeds have variations in traits such as coat color, tail length and ear shape as well as genetic diseases. In this project, a high quality chromosome-scale assembly of American Shorthair breed is established, aiming to uncover the genetic effects on the disease and traits of cats deeper. [Less...](#)

Accession	PRJDB9879
Data Type	Genome sequencing
Scope	Monoisolate
Organism	Felis catus [Taxonomy ID: 9685] Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Carnivora; Feliformia; Felidae; Felinae; Felis; Felis catus
Submission	Registration date: 21-May-2020 NIG
Locus Tag Prefix	AnAms

Project Data:

Resource Name	Number of Links
SEQUENCE DATA	
Nucleotide (Genomic DNA)	20
SRA Experiments	31
OTHER DATASETS	
BioSample	3
Assembly	1
SUPPLEMENTARY DATA	
BioNano map	1

Assembly details:

Assembly	Level	Chrs	BioSample	Taxonomy
GCA_013340865.2	Chromosome	19	SAMD00228405	Felis catus

SRA Data Details

Parameter	Value
Data volume, Gbases	630
Data volume, Tbytes	0.22

See Genome Information for *Felis catus*

NAVIGATE UP

This project is a component of the Kazusa Genome Project: genome sequencing and assembly of Viridiplantae

NAVIGATE ACROSS

147 additional projects are related by organism.

85 additional projects are components of the Kazusa Genome Project: genome sequencing and assembly of Viridiplantae.

関連する BioSample

BioSample

BioSample

Advanced

Search

Help

Organism

Customize ...

Attribute name

breed

collection date

geographic location

isolate

sex

tissue

Customize ...

Sample type

MIGS.eu (1)

Access

Public (3)

Other

Used by SRA (3)

Clear all

Show additional filters

Summary

Sort by Default order

Send to

Filters: [Manage Filters](#)

Links from BioProject

Items: 3

☐ [Domestic cat \(Felis catus\) American shorthair breed \(Takae\) female reproductive organs](#)

1. Identifiers: BioSample: SAMD00635099; SRA: DRS430744

Organism: Felis catus

isolate: Takae; breed: American shorthair

Accession: SAMD00635099 ID: 36884840

[BioProject](#) [SRA](#)

☐ [Domestic cat \(Felis catus\) American shorthair breed \(Senzu\) female reproductive organs](#)

2. Identifiers: BioSample: SAMD00635098; SRA: DRS430743

Organism: Felis catus

isolate: Senzu; breed: American shorthair

Accession: SAMD00635098 ID: 36884839

[BioProject](#) [SRA](#)

☐ [Felis catus American Shorthair breed](#)

3. Identifiers: BioSample: SAMD00228405; SRA: DRS430742

Organism: Felis catus

isolate: Senzu; breed: American Shorthair

Package: MIGS: eukaryote; version 6.0

Accession: SAMD00228405 ID: 15001096

[BioProject](#) [SRA](#) [Nucleotide](#)

Find related data

Database: Select

Find items

Recent activity

Turn Off

Clear

Search

BioSample for BioProject (Select 634486)

(3)

BioSample

File

Felis catus strain:American Shorthair

BioProject

See more...

BioSample: ゲノムサンプル

Felis catus American Shorthair breed																																								
Identifiers	BioSample: SAMD00228405; SRA: DRS430742																																							
Organism	Felis catus (domestic cat) cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Dipnotetrapodomorpha; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Boreoeutheria; Laurasiatheria; Carnivora; Feliformia; Felidae; Felinae; Felis																																							
Package	MIGS: eukaryote, version 6.0																																							
Attributes	<table><tr><td>sample name</td><td>Felis catus American Shorthair breed</td></tr><tr><td>biomaterial provider</td><td>Anicom Specialty Medical Institute Inc.</td></tr><tr><td>breed</td><td>American Shorthair</td></tr><tr><td>collection date</td><td>2018</td></tr><tr><td>broad-scale environmental context</td><td>not applicable</td></tr><tr><td>local-scale environmental context</td><td>not applicable</td></tr><tr><td>environmental medium</td><td>not applicable</td></tr><tr><td>estimated size</td><td>2.44G</td></tr><tr><td>geographic location</td><td>not applicable</td></tr><tr><td>isolation and growth condition</td><td>not applicable</td></tr><tr><td>isolate</td><td>Senzu</td></tr><tr><td>latitude and longitude</td><td>not applicable</td></tr><tr><td>number of replicons</td><td>19</td></tr><tr><td>ploidy</td><td>diploid</td></tr><tr><td>project name</td><td>Genome sequencing project of American Shorthair cat</td></tr><tr><td>propagation</td><td>sexual</td></tr><tr><td>sex</td><td>female</td></tr><tr><td>tissue</td><td>blood</td></tr><tr><td>locus_tag_prefix</td><td>AnAms</td></tr></table>		sample name	Felis catus American Shorthair breed	biomaterial provider	Anicom Specialty Medical Institute Inc.	breed	American Shorthair	collection date	2018	broad-scale environmental context	not applicable	local-scale environmental context	not applicable	environmental medium	not applicable	estimated size	2.44G	geographic location	not applicable	isolation and growth condition	not applicable	isolate	Senzu	latitude and longitude	not applicable	number of replicons	19	ploidy	diploid	project name	Genome sequencing project of American Shorthair cat	propagation	sexual	sex	female	tissue	blood	locus_tag_prefix	AnAms
sample name	Felis catus American Shorthair breed																																							
biomaterial provider	Anicom Specialty Medical Institute Inc.																																							
breed	American Shorthair																																							
collection date	2018																																							
broad-scale environmental context	not applicable																																							
local-scale environmental context	not applicable																																							
environmental medium	not applicable																																							
estimated size	2.44G																																							
geographic location	not applicable																																							
isolation and growth condition	not applicable																																							
isolate	Senzu																																							
latitude and longitude	not applicable																																							
number of replicons	19																																							
ploidy	diploid																																							
project name	Genome sequencing project of American Shorthair cat																																							
propagation	sexual																																							
sex	female																																							
tissue	blood																																							
locus_tag_prefix	AnAms																																							
Description	Whole-genome sequencing sample for the domestic cat (Felis catus) American shorthair Keywords: GSC:MixS;MIGS:6.0																																							
BioProject	PRJDB9879 Felis catus strain:American Shorthair Retrieve all samples from this project																																							

アクセッション番号

生物分類 (NCBI Taxonomy)

パッケージ

属性

[BioProject に戻る](#)

Felis catus strain:American Shorthair (domestic cat) Accession: PRJDB9879 ID: 63448

Genome sequencing project of American Shorthair cat

The domestic cat (*Felis catus*) is one of the most popular companion animals in the world. Modern cat breeds have variations in traits such as coat color, tail length and ear shape as well as genetic diseases. In this project, a high quality chromosome-scale assembly of American Shorthair breed is established, aiming to uncover the genetic effects on the disease and traits of cats deeper. [Less...](#)

Accession	PRJDB9879
Data Type	Genome sequencing
Scope	Monoisolate
Organism	Felis catus [Taxonomy ID: 9685] Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Carnivora; Feliformia; Felidae; Felinae; Felis; Felis catus
Submission	Registration date: 21-May-2020 NIG
Locus Tag Prefix	AnAms

Project Data:

Resource Name	Number of Links
SEQUENCE DATA	
Nucleotide (Genomic DNA)	20
SRA Experiments	31
OTHER DATASETS	
BioSample	3
Assembly	1
SUPPLEMENTARY DATA	
BioNano map	1

▼ Assembly details:

Assembly	Level	Chrs	BioSample	Taxonomy
GCA_013340865.2	Chromosome	19	SAMD00228405	Felis catus

Download

▼ SRA Data Details

Parameter	Value
Data volume, Gbases	630
Data volume, Tbytes	0.22

See Genome Information for *Felis catus*

NAVIGATE UP

This project is a component of the Kazusa Genome Project: genome sequencing and assembly of Viridiplantae

NAVIGATE ACROSS

147 additional projects are related by organism.

85 additional projects are components of the Kazusa Genome Project: genome sequencing and assembly of Viridiplantae.

関連する SRA Experiment

SRA

SRA

Search

Advanced

Help

Access

Public (31)

Source

DNA (29)

RNA (2)

Type

genome (29)

Library Layout

paired (1)

single (30)

Platform

Illumina (1)

PacBio SMRT (30)

Strategy

Genome (29)

other (2)

Data in Cloud

GS (31)

S3 (31)

Clear all

Show additional filters

Summary

20 per page

Send to:

Filters: [Manage Filters](#)

View results as an expanded interactive table using the RunSelector. [Send results to Run selector](#)

Links from BioProject

Items: 1 to 20 of 31

<< First < Prev Page 1 of 2 Next > Last >>

☐ [Sequel sequencing of Domestic cat \(Felis catus\) American Shorthair breed Takae Iso-Seq](#)

1. 1 PACBIO_SMRT (Sequel) run: 21M spots, 30.7G bases, 7.3Gb downloads

Accession: DRX484402

☐ [Sequel sequencing of Domestic cat \(Felis catus\) American Shorthair breed Senzu Iso-Seq](#)

2. 1 PACBIO_SMRT (Sequel) run: 13.4M spots, 20.3G bases, 4.8Gb downloads

Accession: DRX484401

☐ [Sequel sequencing of Domestic cat \(Felis catus\) American Shorthair breed #28 WGS](#)

3. 1 PACBIO_SMRT (Sequel) run: 657,783 spots, 12.1G bases, 2.8Gb downloads

Accession: DRX484400

☐ [Sequel sequencing of Domestic cat \(Felis catus\) American Shorthair breed #27 WGS](#)

4. 1 PACBIO_SMRT (Sequel) run: 690,770 spots, 11.8G bases, 2.8Gb downloads

Accession: DRX484399

☐ [Sequel sequencing of Domestic cat \(Felis catus\) American Shorthair breed #26 WGS](#)

5. 1 PACBIO_SMRT (Sequel) run: 708,339 spots, 12.5G bases, 2.9Gb downloads

Accession: DRX484398

☐ [Sequel sequencing of Domestic cat \(Felis catus\) American Shorthair breed #01 WGS](#)

30. 1 PACBIO_SMRT (Sequel) run: 739,066 spots, 11.8G bases, 2.8Gb downloads

Accession: DRX484373

☐ [Illumina HiSeqX paired end sequencing of Domestic cat \(Felis catus\) American Shorthair breed Senzu WGS](#)

31. 1 ILLUMINA (HiSeq X Five) run: 772.5M spots, 233.3G bases, 111.9Gb downloads

Accession: DRX484372

Find related data

Database: Select

Find items

Recent activity

Turn Off Clear

SRA Links for BioProject (Select 634486) (31) SRA

Felis catus Senzu DNA, unplaced contig, American Shorthair breed Nucleotide

BioSample for BioProject (Select 634486) (3) BioSample

634486[BioProject] (20) Nucleotide

Felis catus strain:American Shorthair BioProject

See more...

13

SRA Experiment: Illumina WGS Paired

DRX484372: Illumina HiSeqX paired end sequencing of Domestic cat (*Felis catus*) American Shorthair

breed Senzu WGS

1 ILLUMINA (HiSeq X Five) run: 772.5M spots, 233.3G bases, 111.9Gb downloads

Submitted by: ANICOM

Study: Genome sequencing project of American Shorthair cat

[PRJDB9879](#) • [DRP012388](#) • [All experiments](#) • [All runs](#)

[show Abstract](#)

Sample: *Felis catus* American Shorthair breed

[SAMD00228405](#) • [DRS430742](#) • [All experiments](#) • [All runs](#)

Organism: [Felis catus](#)

Library:

Name: CatSenzu

Instrument: HiSeq X Five

Strategy: WGS

Source: GENOMIC

Selection: RANDOM

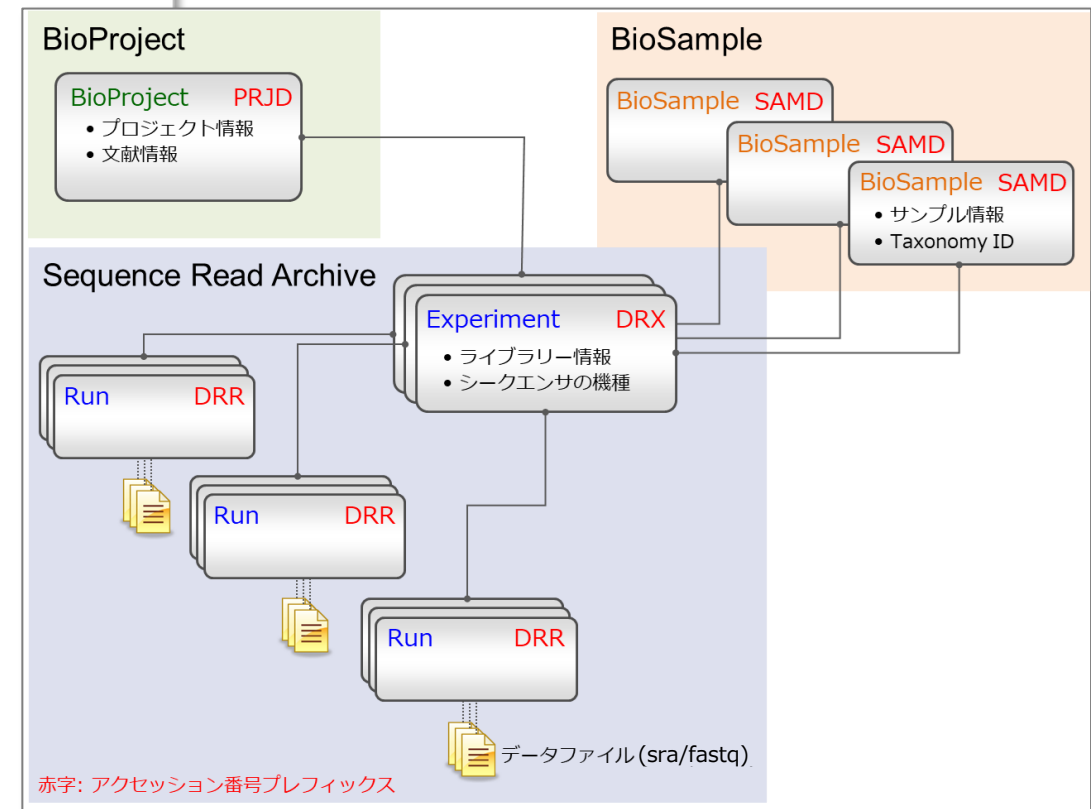
Layout: PAIRED

Construction protocol: DNeasy Blood & Tissue Kit (QIAGEN), TruSeq DNA PCR-Free (Illumina)

Pipeline: [show...](#)

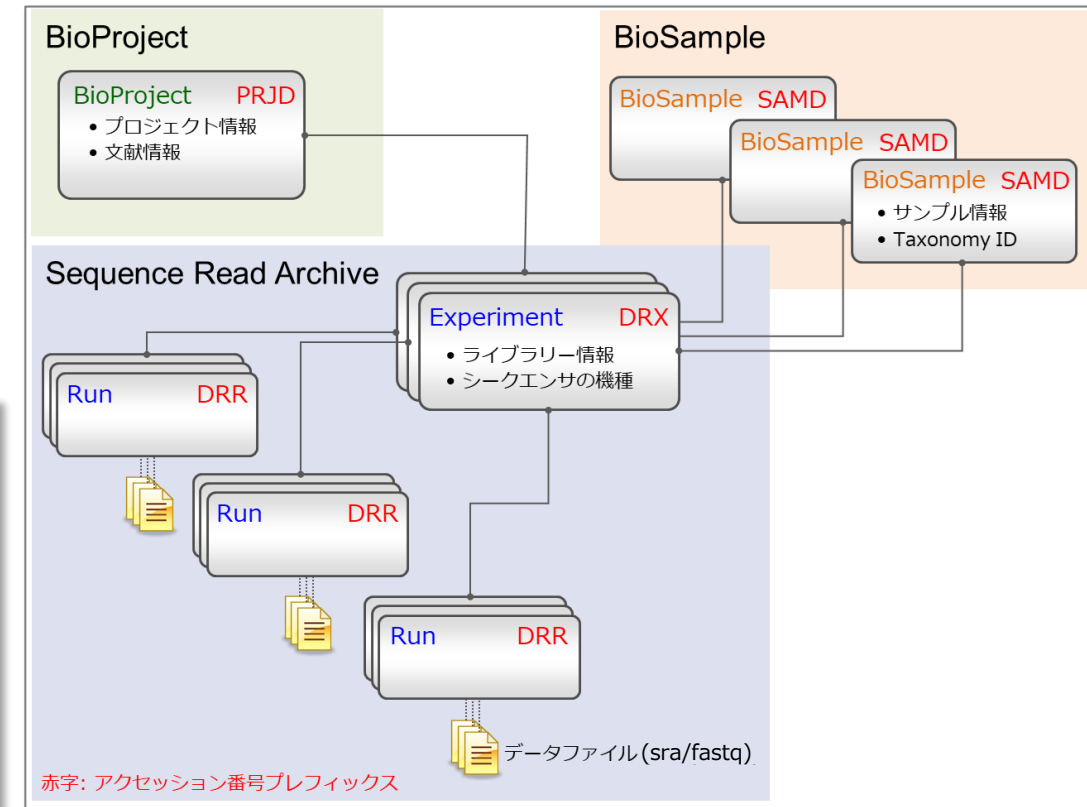
Runs: 1 run, 772.5M spots, 233.3G bases, [111.9Gb](#)

Run	# of Spots	# of Bases	Size	Published
DRR500229	772,491,384	233.3G	111.9Gb	2024-12-24



SRA Run: Reads and Quality Values

Illumina HiseqX paired end sequencing of Domestic cat (*Felis catus*) American Shorthair breed Senzu WGS (DRR500229)

[illegible]

SRA Run: sra/fastq ファイル

Illumina HiseqX paired end sequencing of Domestic cat (Felis catus) American Shorthair breed Senzu WGS (DRR500229)

Metadata Analysis Reads **Data access** FASTA/FASTQ download

SRA archive data

SRA archive data is normalized by the SRA load process and used by the [SRA Toolkit](#) to read and produce formats like FASTQ, SAM, etc. The default toolkit configuration enables it to find and retrieve SRA runs by accession.

Public SRA files are now available from GCP and AWS cloud platforms as well as from NCBI. Access to most data in the cloud requires a user account with the cloud service provider. The user's account will incur costs for cloud compute or to copy data outside of the specified cloud service region.

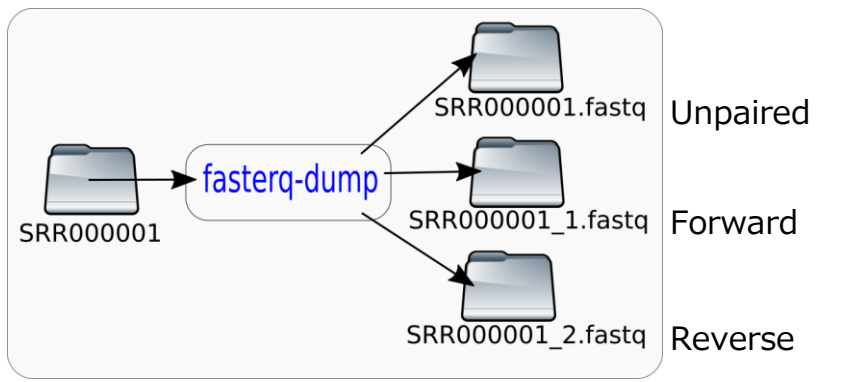
Type	Version	Created	Size	Location	Name
SRA Normalized	1	2024-12-24	111.9GB	AWS	https://sra-pub-run-odp.s3.amazonaws.com/sra/DRR500229/DRR500229
SRA Lite	1	2024-12-24	67.7GB	NCBI	https://sra-downloadb.be-md.ncbi.nlm.nih.gov/sos7/sra-pub-zq-41/DRR000/500/DRR500229/DRR500229.lite.1
				AWS	s3://sra-pub-zq-6/DRR500229/DRR500229.lite.1
				GCP	gs://sra-pub-zq-103/DRR500229/DRR500229.lite.1

Metadata Analysis Reads **Data access** **FASTA/FASTQ download**

Download for Experiment DRX484372

☐ Accession	Total Bases	Spots	
		Total	Filtered
DRR500229	233.3Gbases	772.5M	

This run exceeds the download limit (>5 Gbases). Use [SRA Toolkit](#) to download runs locally in your preferred format.



```
prefetch DRR500229
fasterq-dump DRR500229 --split-files
```

Experiment リストを表形式で閲覧

SRA

SRA

Search

Advanced

Help

Access

Public (31)

Source

DNA (29)

RNA (2)

Type

genome (29)

Library Layout

paired (1)

single (30)

Platform

Illumina (1)

PacBio SMRT (30)

Strategy

Genome (29)

other (2)

Data in Cloud

GS (31)

S3 (31)

Clear all

Show additional filters

Summary

20 per page

Send to:

Filters: [Manage Filters](#)

View results as an expanded interactive table using the RunSelector. [Send results to Run selector](#)

Links from BioProject

Items: 1 to 20 of 31

<< First < Prev Page 1 of 2 Next > Last >>

☐

[Sequel sequencing of Domestic cat \(Felis catus\) American Shorthair breed Takae Iso-Seq](#)

1. 1 PACBIO_SMRT (Sequel) run: 21M spots, 30.7G bases, 7.3Gb downloads

Accession: DRX484402

☐

[Sequel sequencing of Domestic cat \(Felis catus\) American Shorthair breed Senzu Iso-Seq](#)

2. 1 PACBIO_SMRT (Sequel) run: 13.4M spots, 20.3G bases, 4.8Gb downloads

Accession: DRX484401

☐

[Sequel sequencing of Domestic cat \(Felis catus\) American Shorthair breed #28 WGS](#)

3. 1 PACBIO_SMRT (Sequel) run: 657,783 spots, 12.1G bases, 2.8Gb downloads

Accession: DRX484400

☐

[Sequel sequencing of Domestic cat \(Felis catus\) American Shorthair breed #27 WGS](#)

4. 1 PACBIO_SMRT (Sequel) run: 690,770 spots, 11.8G bases, 2.8Gb downloads

Accession: DRX484399

☐

[Sequel sequencing of Domestic cat \(Felis catus\) American Shorthair breed #26 WGS](#)

5. 1 PACBIO_SMRT (Sequel) run: 708,339 spots, 12.5G bases, 2.9Gb downloads

Accession: DRX484398

☐

[Sequel sequencing of Domestic cat \(Felis catus\) American Shorthair breed #01 WGS](#)

30. 1 PACBIO_SMRT (Sequel) run: 739,066 spots, 11.8G bases, 2.8Gb downloads

Accession: DRX484373

☐

[Illumina HiSeqX paired end sequencing of Domestic cat \(Felis catus\) American Shorthair breed Senzu WGS](#)

31. 1 ILLUMINA (HiSeq X Five) run: 772.5M spots, 233.3G bases, 111.9Gb downloads

Accession: DRX484372

Find related data

Database: Select

Find items

Recent activity

Turn Off Clear

Q

SRA Links for BioProject (Select 634486) (31)

SRA

Q

Felis catus Senzu DNA, unplaced contig, American Shorthair breed

Nucleotide

Q

BioSample for BioProject (Select 634486) (3)

BioSample

Q

634486[BioProject] (20)

Nucleotide

Q

Felis catus strain:American Shorthair

BioProject

See more...

17

SRA Run Selector: 表形式で閲覧

Filters List

1

☐

Assay Type

2

☐

AvgSpotLen

3

☐

Bases

4

☐

BioSample

5

☐

Bytes

6

☐

Collection_Date

7

☐

Instrument

8

☐

LibraryLayout

9

☐

LibrarySelection

10

☐

LibrarySource

11

☐

Platform

12

☐

Sample Name

13

☐

Sample_name

14

☐

geo_loc_name

15

☐

geo_loc_name_country

16

☐

geo_loc_name_country_continent

Common Fields

BioProject	PRJDB9879
Consent	PUBLIC
BREED	American Shorthair
Center Name	ANICOM
DATASTORE filetype	RUN.ZQ, SRA
DATASTORE provider	GS, NCBI, S3
DATASTORE region	gs.us-east1, ncbi.public, s3.us-east-1
Organism	Felis catus
project name	Genome sequencing project of American Shorthair cat

Select

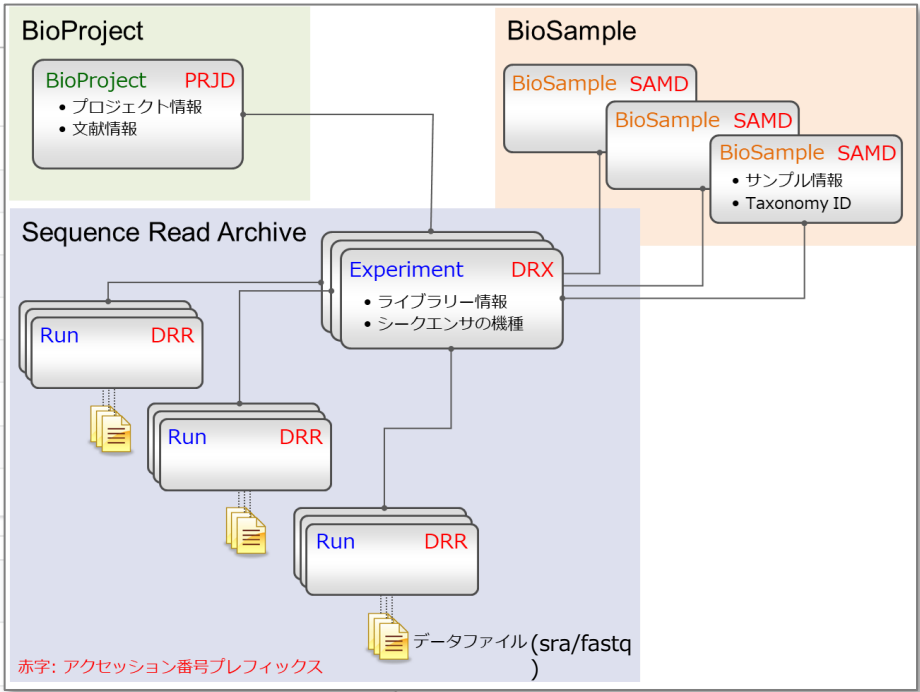
	Runs	Bytes	Bases	Download
Total	31	205.10 Gb	629.70 G	Metadata or Accession List
Selected	0	0	0	Metadata or Accession List or JWT Cart

Found 31 Items

Search within results

Q Clear

☒	☐	Run 1	BioSample 2	Assay Type 3	Collection_Date 4	Experiment 5	Instrument 6	isolate 7	Library Name 8	LibraryLayout 9
☐	1	DRR500229	SAMD00228405	WGS	2018	DRX484372	HiSeq X Five	Senzu	CatSenzu	PAIRED
☐	2	DRR500230	SAMD00228405	WGS	2018	DRX484373	Sequel	Senzu	cat01	SINGLE
☐	3	DRR500231	SAMD00228405	WGS	2018	DRX484374	Sequel	Senzu	cat02	SINGLE
⋮										⋮
☐	29	DRR500257	SAMD00228405	WGS	2018	DRX484400	Sequel	Senzu	cat28	SINGLE
☐	30	DRR500258	SAMD00635098	FL-cDNA	2018-10-31	DRX484401	Sequel	Senzu	senzu_isoseq	SINGLE
☐	31	DRR500259	SAMD00635099	FL-cDNA	2018-09-28	DRX484402	Sequel	Takae	takae_isoseq	SINGLE



BioProject に再び戻る

Felis catus strain:American Shorthair (domestic cat)
Accession: PRJDB9879 ID: 63448

Genome sequencing project of American Shorthair cat

The domestic cat (*Felis catus*) is one of the most popular companion animals in the world. Modern cat breeds have variations in traits such as coat color, tail length and ear shape as well as genetic diseases. In this project, a high quality chromosome-scale assembly of American Shorthair breed is established, aiming to uncover the genetic effects on the disease and traits of cats deeper. [Less...](#)

Accession	PRJDB9879
Data Type	Genome sequencing
Scope	Monoisolate
Organism	Felis catus [Taxonomy ID: 9685] Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Carnivora; Feliformia; Felidae; Felinae; Felis; Felis catus
Submission	Registration date: 21-May-2020 NIG
Locus Tag Prefix	AnAms

See Genome Information for *Felis catus*

NAVIGATE UP

This project is a component of the Kazusa Genome Project: genome sequencing and assembly of Viridiplantae

NAVIGATE ACROSS

147 additional projects are related by organism.

85 additional projects are components of the Kazusa Genome Project: genome sequencing and assembly of Viridiplantae.

Project Data:

Resource Name	Number of Links
SEQUENCE DATA	
Nucleotide (Genomic DNA)	20
SRA Experiments	31
OTHER DATASETS	
BioSample	3
Assembly	1
SUPPLEMENTARY DATA	
BioNano map	1

▼ Assembly details:

Assembly	Level	Chrs	BioSample	Taxonomy
GCA_013340865.2	Chromosome	19	SAMD00228405	Felis catus

[Download](#)

▼ SRA Data Details

Parameter	Value
Data volume, Gbases	630
Data volume, Tbytes	0.22

関連するアセンブリ配列

Nucleotide

Nucleotide

634486[BioProject]

×

Search

Create alert Advanced

Help

Species

Animals (20)

Customize ...

Molecule types

genomic DNA/RNA (20)

Customize ...

Source databases

INSDC (GenBank) (20)

Customize ...

Sequence Type

Nucleotide (20)

Sequence length

Summary

50 per page

Sort by Default order

Send to:

Filters: [Manage Filters](#)

Items: 20

☐ [Felis catus Senzu DNA, chromosome: X, American Shorthair breed](#)

1. 129,629,249 bp linear DNA

Accession: AP023170.1 GI: 1848010751

[Assembly](#) [BioProject](#) [BioSample](#) [Taxonomy](#)

[GenBank](#) [FASTA](#) [Graphics](#)

☐ [Felis catus Senzu DNA, chromosome: F2, American Shorthair breed](#)

2. 86,116,642 bp linear DNA

Accession: AP023169.1 GI: 1848010750

[Assembly](#) [BioProject](#) [BioSample](#) [Taxonomy](#)

[GenBank](#) [FASTA](#) [Graphics](#)

⋮

☐ [Felis catus Senzu DNA, chromosome: A1, American Shorthair breed](#)

19. 243,504,654 bp linear DNA

Accession: AP023152.1 GI: 1848010732

[Assembly](#) [BioProject](#) [BioSample](#) [Taxonomy](#)

[GenBank](#) [FASTA](#) [Graphics](#)

☐ [Felis catus Senzu DNA, unplaced contig, American Shorthair breed](#)

20. 1,381 bp linear DNA

Accession: AP023171.1 GI: 1843987454

[Assembly](#) [BioProject](#) [BioSample](#) [Taxonomy](#)

[GenBank](#) [FASTA](#) [Graphics](#)

Analyze these sequences

Run BLAST

Find related data

Database:

Select

Find items

Search details

634486[BioProject]

20

DDBJ Search

[DDBJ Search Home](#) > [Entries](#) > [BioProject](#) > [PRJDB9879](#)

[About DDBJ Search](#) [API](#)

Title	Genome sequencing project of American Shorthair cat
Name	Felis catus strain:American Shorthair
Description	The domestic cat (<i>Felis catus</i>) is one of the most popular companion animals in the world. Modern cat breeds have variations in traits such as coat color, tail length and ear shape as well as genetic diseases. In this project, a high quality chromosome-scale assembly of American Shorthair breed is established, aiming to uncover the genetic effects on the disease and traits of cats deeper.
Organism	Felis catus 
Organizations	Role: owner / Type: center NIG
Project Type	Genome sequencing

Type	BioProject
Accession	PRJDB9879
Status 	Public
Accessibility 	Public Access
Date Submitted	2020-05-13
Date First Published	2020-05-22
Date Last Published	2020-05-22

Properties **Open Properties**

Related Links

BioSample (3)	SAMD00228405	SAMD00635098	SAMD00635099			
SRA Run (31)	DRR500229	DRR500230	DRR500231	DRR500232	DRR500233	DRR500234
	DRR500235	DRR500236	DRR500237	DRR500238	DRR500239	DRR500240
	DRR500241	DRR500242	DRR500243	DRR500244	DRR500245	DRR500246
	DRR500247	DRR500248	DRR500249	DRR500250	DRR500251	DRR500252
	DRR500253	DRR500254	DRR500255	DRR500256	DRR500257	DRR500258
	DRR500259					
SRA Experiment (31)	DRX484372	DRX484373	DRX484374	DRX484375	DRX484376	DRX484377
	DRX484378	DRX484379	DRX484380	DRX484381	DRX484382	DRX484383
	DRX484384	DRX484385	DRX484386	DRX484387	DRX484388	DRX484389
	DRX484390	DRX484391	DRX484392	DRX484393	DRX484394	DRX484395
	DRX484396	DRX484397	DRX484398	DRX484399	DRX484400	DRX484401
	DRX484402					
SRA Study (1)	DRP012388					
insdc (20)	AP023152 	AP023153 	AP023154 	AP023155 	AP023156 	AP023157 
	AP023158 	AP023159 	AP023160 	AP023161 	AP023162 	AP023163 
	AP023164 	AP023165 	AP023166 	AP023167 	AP023168 	AP023169 
	AP023170 	AP023171 				
insdc-assembly (1)	GCA_013340865 					

Download

JSON

JSON-LD

INSDC 構成データベース

Data type	Research Organization of Information and Systems (ROIS), National Institute of Genetics (NIG)	European Molecular Biology Laboratory (EMBL), European Bioinformatics Institute (EBI)	National Library of Medicine (NLM), National Center for Biotechnology Information (NCBI) at the National Institutes of Health
Next Generation reads	Sequence Read Archive	European Nucleotide Archive	Sequence Read Archive
Assembled Sequences	DDBJ		GenBank
Samples	BioSample		BioSample
Studies	BioProject		BioProject

SRA - NGS リード

- リードは Run 単位のファイル
- メタデータとリードを分離

DDBJ/ENA/GenBank - アセンブリ配列

- 1配列1ファイル
- メタデータと配列は一体

データを登録する

BioProject

BioSample

DRA (DDBJ SRA)

DDBJ

登録システム

データベース	登録システム	対象	アカウント
BioProject	D-way	プロジェクト情報	DDBJ アカウント
BioSample		サンプル情報	
DRA		NGS リード	
DDBJ	NSSS: ウェブ塩基配列登録システム	配列数100以下、かつ、配列長500kb以下	アカウント不要
	MSS: ファイルベースの登録システム	上記基準以上の大規模配列（ゲノム等）	DDBJ アカウント

D-way 登録システム

D-way

DDBJ Submission Portal D-way

(August 14 to 17) DDBJ Center Closing for Summer Break

Revision of the Terms of Use (Access and Benefit-Sharing)

Login to submit data to DDBJ BioProject, BioSample, Sequence Read Archive (DRA) and Genomic Expression Archive (GEA).

Login

Login ID:

Password:

Login

- Never submit data without the permission of the principal investigator.
研究責任者の許可なくデータを登録しないでください。
- When you are submitting data derived from human subjects, please read "[Submission of research data from human subjects](#)".
ヒトに由来するデータを登録する場合は「ヒトを対象とした研究データの登録について」を熟読してください。

Forgot password ([English](#), [Japanese](#))?
Forgot username ([English](#), [Japanese](#))?
[Register for a new account](#)

<https://trace.ddbj.nig.ac.jp/D-way/>

DDBJ アカウント

DDBJ Account

New Account

Username *

Must be 3–24 characters, start with a lowercase letter or digit (not _ or -), and contain only lowercase letters, digits, underscores, or hyphens.

Password *

At least 8 characters.

Password confirmation *

Email *

First name *

First name in Japanese

Middle name

Last name *

Last name in Japanese

☐ I have read and agree to the [Terms of Use](#). *

Create Account

DDBJ アカウント新規作成

 **DDBJ**
DNA Data Bank of Japan

Sign in to your account

Username or email

Password

👁

[Forgot Password?](#)

Sign In

Profile

Password

SSH keys

Logout

SSH keys

New SSH key

No SSH keys

ログイン後、SSH 公開鍵を登録
データファイルの sftp アップロードに必要
<https://www.ddbj.nig.ac.jp/ddbj-account.html>

27

D-way ログイン



DDBJ Submission Portal D-way

(August 14 to 17) DDBJ Center Closing for Summer Break

Revision of the Terms of Use (Access and Benefit-Sharing)

Login to submit data to DDBJ BioProject, BioSample, Sequence Read Archive (DRA) and Genomic Expression Archive (GEA).

Login

Login ID:	test09
Password:	••••••••
	Login

- Never submit data without the permission of the principal investigator.
研究責任者の許可なくデータを登録しないでください。
- When you are submitting data derived from human subjects, please read "[Submission of research data from human subjects](#)".
ヒトに由来するデータを登録する場合は「[ヒトを対象とした研究データの登録について](#)」を熟読してください。

Forgot password ([English](#), [Japanese](#))?

Forgot username ([English](#), [Japanese](#))?

[Register for a new account](#)

[\(August 14 to 17\) DDBJ Center Closing for Summer Break](#)

[Revision of the Terms of Use \(Access and Benefit-Sharing\)](#)

Account: test09

[登録ナビゲーションサイト](#): データに関するいくつかの質問に答えると登録方法が案内されます。

[Submission navigation site](#): By answering several questions about your data, you will be guided how to submit your data.

[Register a public key](#) to submit data to DRA/GEA.

DRA/GEA にデータを登録するためには、[公開鍵の登録](#)が必要です。

When you are submitting data derived from human subjects, please read "[Submission of research data from human subjects](#)".

ヒトに由来するデータを登録する場合は「[ヒトを対象とした研究データの登録について](#)」を熟読してください。

❗ Use only ASCII characters (English letters) throughout your submission. Non-ASCII characters (for example, Japanese and special characters) are not accepted.

登録には ASCII 文字 (英語文字) のみを使用してください。非 ASCII 文字 (日本語や特殊文字など) は受け付けていません。

❗ If there is no reply from submitters after three months of initial contact from us, submissions will be cancelled.

登録者に問い合わせた後三か月以上回答が無い場合は Submission をキャンセルいたします。

[BioProject](#)

A collection of biological data related to a single initiative, originating from a single organization or from a consortium.

You can register a project here and later submit and link sequence data to the project. [Manual](#)

ここからプロジェクトのみを登録し、後から配列データを登録してプロジェクトに関連付けることができます。 [マニュアル](#)

[BioSample](#)

Descriptions of biological source materials used to generate experimental data in any of DDBJ's primary data archives.

You can register sample(s) here and later submit and link sequence data to the sample(s). [Manual](#)

ここからサンプルのみを登録し、後から配列データを登録してサンプルに関連付けることができます。 [マニュアル](#)

[DDBJ Sequence Read Archive \(DRA\)](#)

DRA stores raw and aligned sequence data from next-generation sequencing platforms.

You can complete both your BioProject as well as BioSample submissions within DRA submission interface. [Manual](#)

DRA 登録インターフェースでまとめて BioProject, BioSample と DRA メタデータを登録することができます。 [マニュアル](#)

BioProject 登録: SUBMITTER

BioProject Submission ID: PSUB049081

SUBMITTER

GENERAL INFO

PROJECT TYPE

TARGET

PUBLICATION

OVERVIEW

Submitter

First name : Hanako *Last name : Mishima

*Email : test09@nig.ac.jp

First name : Ichiro *Last name : Fuji

*Email : test10@nig.ac.jp

[+ Add another submitter](#)

Enter submitting organization.

Organization

*Submitting organization : National Institute of Genetics

Submitting organization URL :

Specify when this submission should be released to the public.

Data Release

☒ Hold (not viewable until the release of linked data)

☐ Release immediately following curation

Continue

BioProject 登録: GENERAL INFO

SUBMITTER

GENERAL INFO

PROJECT TYPE

TARGET

PUBLICATION

OVERVIEW

Project Description

*Project title : Genome sequencing project of American Shorthair cat

*Description : The domestic cat (Felis catus) is one of the most popular companion animals in the world. Modern cat breeds have variations in traits such as coat color, tail length and ear shape as well as genetic diseases. In this project, a high quality chromosome-scale assembly of American Shorthair breed is established, aiming to uncover the genetic effects on the disease and traits of cats deeper.

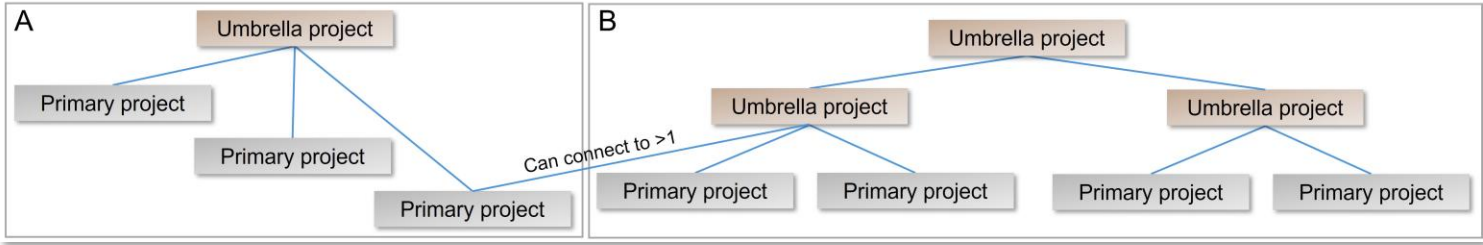
Provide a description (a paragraph) of the study goals and relevance (more than 100 characters).

For umbrella projects, please include the comment: "Please register this as an umbrella project."
For human data submissions approved by DBCLS, please provide both the hum ID and application ID (e.g., hum0001, J-DS000001-001).

Comments

Private comments to DDBJ staff :

アンブレラプロジェクトとしての登録を希望する場合、Comments に記載



BioProject 登録: PROJECT TYPE

SUBMITTERGENERAL INFOPROJECT TYPETARGETPUBLICATIONOVERVIEW

A general label indicating the primary study goal.

*Project data type

Genome Sequencing :☒

Clone Ends :☐

Epigenomics :☐

Exome :☐

Map :☐

Metagenome :☐

Phenotype and Genotype :☐

Proteome :☐

Random Survey :☐

Targeted Locus (Loci):☐

Transcriptome or Gene Expression :☐

Variation :☐

Other :☐

Indicates the scope and purity of the biological sample used for the study.

Sample scope/Material/Capture/Methodology

*Sample scope : Monoisolate

*Material : Genome

*Capture : Whole

*Methodology : Sequencing

Project goals with respect to the type of data that will be generated and submitted to an INSDC-associated database.

*Objective

Raw Sequence Reads :☒

Sequence :☒

Analysis :☐

Assembly :☒

Annotation :☐

Variation :☐

Epigenetic Markers :☐

Expression :☐

Maps :☐

Phenotype :☐

Other :☐

Continue

32

BioProject 登録: TARGET ~ OVERVIEW

SUBMITTER

GENERAL INFO

PROJECT TYPE

TARGET

PUBLICATION

OVERVIEW

If the project targets multiple organisms, enter a common lineage e.g. "Mammalia" (taxonomy id: 40674) for the project targeting "Homo sapiens" and "Mus musculus".
If an organism name is not found in the NCBI Taxonomy, enter a lineage or a species name with sp. e.g. "Escherichia sp." (tax id: 1884818) for a novel species "Escherichia sp. ABCDEF".
Apply a novel organism name in [BioSample](#).

Organism information

*Organism name : Felis catus Taxonomy ID : 9685

Strain, breed, cultivar : OR Isolate name or label :

Continue



SUBMITTER

GENERAL INFO

PROJECT TYPE

TARGET

PUBLICATION

OVERVIEW

Publication

PubMed ID : OR DOI : 10.1016/j.jare.2024.10.023

+ Add another publication

Continue



SUBMITTER

GENERAL INFO

PROJECT TYPE

TARGET

PUBLICATION

OVERVIEW

SUBMITTER

Submitter 1

First name Hanako

Last name Mishima

Email test09@nig.ac.jp

Submitter 2

First name Ichiro

Last name Fuji

Email test10@nig.ac.jp

Organization

Submitting organization National Institute of Genetics

Data Release

Data Release Hold

GENERAL INFO

Project Description

Project title Felis catus strain:American Shorthair (domestic cat)

...

TARGET

Organism Information

Organism name Felis catus

Taxonomy ID 9685

PUBLICATION

Publication 1

PubMed ID

DOI 10.1016/j.jare.2024.10.023

☒ I confirm that, to the best of my knowledge, the data submitted here are not subject to any restrictions that would prevent their open sharing through the databases operated by the BioData Science Initiative (BSI). I have also reviewed and agreed to the BSI Terms of Use, including the provisions on the rights and duties of submitters, the handling of human-derived research data, and the access and benefit-sharing (ABS) of genetic resources.
Read more:

- BSI Terms of Use: <https://www.ddbj.nig.ac.jp/policies-e.html>
- Applicable national and international laws and rules on access and benefit-sharing: <https://www.ddbj.nig.ac.jp/faq/en/abs-data-sharing-e.html>

Submit

利用規約に同意して登録

BioProject: アクセッション番号

BioProject submission list for test09

New submission

⚠ To update an existing record or recent submission, please send your request with your Submission ID or BioProject accession included ([English](#), [Japanese](#)).

BioProject Submission ID ▾	BioProject Accession ⇅	Status ⇅	Title ⇅	Project Type ⇅	Created Date ⇅
PSUB049081	PRJDB42877	Private	Genome sequencing project of American Shorthair cat	Primary	2026-08-03

以降の登録や論文で引用

[\(August 14 to 17\) DDBJ Center Closing for Summer Break](#)

[Revision of the Terms of Use \(Access and Benefit-Sharing\)](#)

Account: test09

[登録ナビゲーションサイト](#): データに関するいくつかの質問に答えると登録方法が案内されます。

[Submission navigation site](#): By answering several questions about your data, you will be guided how to submit your data.

[Register a public key](#) to submit data to DRA/GEA.

DRA/GEA にデータを登録するためには、[公開鍵の登録](#)が必要です。

When you are submitting data derived from human subjects, please read "[Submission of research data from human subjects](#)".

ヒトに由来するデータを登録する場合は「[ヒトを対象とした研究データの登録について](#)」を熟読してください。

❗ Use only ASCII characters (English letters) throughout your submission. Non-ASCII characters (for example, Japanese and special characters) are not accepted.

登録には ASCII 文字 (英語文字) のみを使用してください。非 ASCII 文字 (日本語や特殊文字など) は受け付けていません。

❗ If there is no reply from submitters after three months of initial contact from us, submissions will be cancelled.

登録者に問い合わせた後三か月以上回答が無い場合は Submission をキャンセルいたします。

[BioProject](#)

A collection of biological data related to a single initiative, originating from a single organization or from a consortium.

You can register a project here and later submit and link sequence data to the project. [Manual](#)

ここからプロジェクトのみを登録し、後から配列データを登録してプロジェクトに関連付けることができます。 [マニュアル](#)

[BioSample](#)

Descriptions of biological source materials used to generate experimental data in any of DDBJ's primary data archives.

You can register sample(s) here and later submit and link sequence data to the sample(s). [Manual](#)

ここからサンプルのみを登録し、後から配列データを登録してサンプルに関連付けることができます。 [マニュアル](#)

[DDBJ Sequence Read Archive \(DRA\)](#)

DRA stores raw and aligned sequence data from next-generation sequencing platforms.

You can complete both your BioProject as well as BioSample submissions within DRA submission interface. [Manual](#)

DRA 登録インターフェースでまとめて BioProject, BioSample と DRA メタデータを登録することができます。 [マニュアル](#)

BioSample 登録: SUBMITTER ~ GENERAL INFO

BioSample Submission ID: SSUB048035

SUBMITTER

GENERAL INFO

SAMPLE PACKAGE

ATTRIBUTES

OVERVIEW

Submitter

First name : Hanako *Last name : Mishima

*Email : test09@nig.ac.jp

First name : Ichiro *Last name : Fuji

*Email : test10@nig.ac.jp

+ Add another submitter

Enter submitting organization.

Organization

*Submitting organization : National Institute of Genetics

Submitting organization URL :

Continue



SUBMITTER

GENERAL INFO

SAMPLE PACKAGE

ATTRIBUTES

OVERVIEW

Specify when this submission should be released to the public.

Data Release

☒ Hold (not viewable until the release of linked data)

☐ Release immediately following processing

External Links

Link description URL

+ Add another link

Comments

Private comments to DDBJ staff :

Continue

BioSample 登録: SAMPLE PACKAGE

SUBMITTER

GENERAL INFO

SAMPLE PACKAGE

ATTRIBUTES

OVERVIEW

Select the package that best describes your samples.
For more information, please see the sample type page ([English](#), [Japanese](#)).

Sample package

*Core Package :

- ☒ Standard
Use when it is not advantageous to use the Genome Standards Consortium (GSC) minimum information (MIxS) packages.
- ☐ SARS-CoV-2: clinical or host-associated
- ☐ SARS-CoV-2: wastewater surveillance
- ☐ Microbe
- ☐ Model organism or animal
- ☐ Metagenome or environmental
- ☐ Invertebrate
- ☐ Human
- ☐ Plant
- ☐ Virus
- ☐ Beta-lactamase
- ☐ Omics
- ☐ Pathogen
- ☐ Genome, metagenome or marker sequences (MIxS compliant)

Continue

サンプルパッケージ

DDBJ/NCBI で定義

Standard
SARS-CoV-2: clinical or host-associated
SARS-CoV-2: wastewater surveillance
Microbe
Model organism or animal
Metagenome or environmental
Invertebrate
Human
Plant
Virus
Beta-lactamase
Omics

生物種
データ種別

Pathogen
Pathogen: clinical or host-associated
Pathogen: environmental/food/other

Genomic Standards Consortium (GSC) で定義

MixS
MIGS: cultured bacteria/archaea
MIGS: eukaryote
MIGS: virus
MIMS: metagenome/environmental
MIMAG: metagenome-assembled genome
MISAG: single amplified genome
MIMARKS: specimen
MIMARKS: survey
MIUVIG: uncultivated virus genome

データ種別



Environmental Package	
No environmental package	human-skin
agriculture	human-vaginal
air	hydrocarbon-cores
built	hydrocarbon-fluids_swabs
food-animal	microbial
food-farm_env	miscellaneous
food-human_foods	plant-associated
food-prod_facility	sediment
host-associated	soil
human-associated	symbiont-associated
human-gut	wastewater
human-oral	water

サンプリング環境

生物種やデータ種別に応じた必須・任意のサンプル属性セットを提供する仕組み
サンプル採取地（geo_loc_name）と採取日（collection_date）は全パッケージで必須

BioSample 登録: SAMPLE PACKAGE – MIGS.eu

SUBMITTERGENERAL INFOSAMPLE PACKAGEATTRIBUTESOVERVIEW

Select the package that best describes your samples.
For more information, please see the sample type page ([English](#), [Japanese](#)).

Sample package

*Core Package :

☒ Genome, metagenome or marker sequences (MIXS compliant)

Use for genomes, metagenomes, and marker sequences. These samples include specific attributes that have been defined by the Genome Standards Consortium (GSC).

☐ Cultured Bacterial/Archaeal Genomic Sequences (MIGS)

☒ Eukaryotic Genomic Sequences (MIGS)

☐ Viral Genomic Sequences (MIGS)

☐ Environmental/Metagenome Genomic Sequences (MIMS)

☐ Metagenome-assembled Genome Sequences (MIMAG)

☐ Single Amplified Genome Sequences (MISAG)

☐ Specimen Marker Sequences (MIMARKS)

☐ Survey-related Marker Sequences (MIMARKS)

☐ Uncultivated Viral Genome Sequences (MIUVIG)

Environmental package (MIXS Sample)

*Environmental package (MIXS Sample) :

☒ No environmental package

☐ Agriculture

☐ Air

☐ Built

☐ Food-animal and animal feed

☐ Food-farm environment

☐ Food-food production facility

☐ Food-human foods

☐ Host-associated

☐ Human-associated

☐ Human-gut

☐ Human-oral

☐ Human-skin

☐ Human-vaginal

☐ Hydrocarbon resources-cores

☐ Hydrocarbon resources-fluids/swabs

☐ Microbial mat/biofilm

☐ Miscellaneous or artificial

☐ Plant-associated

☐ Sediment

☐ Soil

☐ Symbiont-associated

☐ Wastewater/sludge

☐ Water

Continue

BioSample 登録: ATTRIBUTES

SUBMITTER GENERAL INFO SAMPLE PACKAGE ATTRIBUTES OVERVIEW

Describe your sample(s) by providing sample attributes.
For more information, please see the sample attribute page ([English](#), [Japanese](#))

Attributes

*Attributes file : ファイルを選択 選択されていません

[Download a template text file for BioSample attributes](#)

Download and edit the template text file in spreadsheet or text editor.

- You can submit multiple samples described in lines.
サンプルを複数行に記載し、まとめて登録することができます。
- If you do not have information for the required attribute(s), please provide the value as either 'not collected', 'not applicable' or 'missing'.
必須属性に対する値がない場合は、'not collected', 'not applicable' もしくは 'missing' を記入してください。
- Provide 'organism' and 'taxonomy_id' as registered in [NCBI Taxonomy \(for unregistered organism\)](#)
[NCBI Taxonomy](#)に登録されている生物を、'organism' と 'taxonomy_id' に記入してください ([未登録の生物の場合](#))

Continue

*sample_name	*sample_title	description
Felis catus American Shorthair breed	Felis catus American Shorthair breed	Whole-genome sequencing sample for the domestic cat (Felis catus) American shorthair

*organism	taxonomy_id	bioproject_id	locus_tag_prefix	strain	isolate	cultivar	ecotype
Felis catus		PRJDB9879			Senzu		

*collection_date	*env_broad_scale	*env_local_scale	*env_medium	*geo_loc_name	...	sex	tissue
2018	missing	missing	missing	Japan	...	femal	blood

タブ区切りテキストファイルにサンプル属性を記載（1行1サンプルで複数サンプルを登録可）

BioSample 登録: INSDC missing value reporting terms

INSDC term (top level)	INSDC term (lower level)	Definition	INSDC term (reporting level)
not applicable		Information is inappropriate to report, can indicate that the standard itself fails to model or represent the information appropriately	control sample
			sample group
missing	not collected	Information of an expected format was not given because it has not been collected	synthetic construct
			lab stock
			third party data
	not provided	Information of an expected format was not given, a value may be given at the later stage	data agreement established pre-2023
	restricted access	Information exists but can not be released openly because of privacy concerns	endangered species
			human-identifiable

必須属性値を記入できない場合の記載方法

- geo_loc_name と collection_date - missing: reporting level term (例 missing: lab stock) を記載
- これら二属性以外は top と lower level terms も記載可 (例 missing)

BioSample 登録: ATTRIBUTES - validation

✖ 1 error

⚠ 1 warning

SUBMITTER

GENERAL INFO

SAMPLE PACKAGE

ATTRIBUTES

OVERVIEW

Error Message

error

Attribute value is not in controlled terms. (explanation: [English](#), [Japanese](#))

Show 10 entries

Sample name	Attribute	Attribute value
Felis catus American Shorthair breed	sex	femal

Showing 1 to 1 of 1 entries

warning

If applicable, the taxonomy id will be automatically filled and the organism scientific name. When the organism(s) is novel, please enter proposed name(s) in the taxonomy id empty and submit the BioSample. (explanation: [English](#), [Japanese](#))

Show 10 entries

Sample name	organism	taxonomy_id	Suggested value (taxonomy_id)
Felis catus American Shorthair breed	Felis catus		9685

Showing 1 to 1 of 1 entries

Previous1Next

Describe your sample(s) by providing sample attributes.
For more information, please see the sample attribute page ([English](#), [Japanese](#))

Attributes

MIGS_eu.txtDelete

Continue

BS_R0138 error Invalid Attribute value for controlled terms
message: Attribute value is not in controlled terms [?] .

attribute	controlled terms
rel_to_oxygen	"aerobe", "anaerobe", "facultative", "microaerophilic", "microanaerobe", "obligate"
sediment_type	"biogenous", "cosmogenous", "hydrogenous", "lithogenous"
sex	"male", "female", "pooled male and female", "neuter", "hermaphrodite", "intersex"
space_typ_state	"typical occupied", "typically unoccupied", "missing", "not applicable", "not collect"

エラーがなくなるようにファイルを修正

BioSample 登録: OVERVIEW

SUBMITTERGENERAL INFOSAMPLE PACKAGEATTRIBUTESOVERVIEW

SUBMITTER

Submitter 1

First name	Mishima
Last name	Hanako
Email	test09@nig.ac.jp

Submitter 2

First name	Ichiro
Last name	Fuji
Email	test10@nig.ac.jp

Organization

Submitting organization	National Institute of Genetics
Submitting organization URL	

GENERAL INFO

Data Release

Data Release	Hold
--------------	------

External Links 1

Link description	
URL	

Private comments to DDBJ staff

Comments	
----------	--

SAMPLE PACKAGE

Core Package	MiXs compliant
MiXs Sample	Eukaryotic Genomic Sequences (MIGS)
MiXs Sample	No environmental package

ATTRIBUTES

Attributes

TSV file	SSUB048035.txt
----------	----------------

☒ I confirm that, to the best of my knowledge, the data submitted here are not subject to any restrictions that would prevent their open sharing through the databases operated by the BioData Science Initiative (BSI). I have also reviewed and agreed to the BSI Terms of Use, including the provisions on the rights and duties of submitters, the handling of human-derived research data, and the access and benefit-sharing (ABS) of genetic resources.

Read more:

BSI Terms of Use: <https://www.ddbj.nig.ac.jp/policies-e.html>

Applicable national and international laws and rules on access and benefit-sharing: <https://www.ddbj.nig.ac.jp/faq/en/abs-data-sharing-e.html>

Submit

43

BioSample 登録: アクセッション番号

BioSample Submission ID: SSUB048035

SUBMITTERGENERAL INFOSAMPLE PACKAGEATTRIBUTESOVERVIEW

Sample Summary

Sample Name	Status	BioSample Accession	Published Date	Modified Date
Felis catus American Shorthair breed	Private	SAMD01949766	-----	2026-08-04

SUBMITTER

Submitter 1

First name	Hanako
Last name	Mishima
Email	test09@nig.ac.jp

Organization

Submitting organization	National Institute of Genetics
Submitting organization URL	

GENERAL INFO

Data Release

Data Release	Hold
--------------	------

External Links 1

Link description	
URL	

Private comments to DDBJ staff

Comments	
----------	--

SAMPLE PACKAGE

Core Package	MiXS compliant
MiXS Sample	Eukaryotic Genomic Sequences (MIGS)
MiXS Sample	No environmental package

ATTRIBUTES

Attributes

TSV file	SSUB048035.txt
----------	----------------

← SAMD01949766
以降の登録と論文で引用

← テキストファイルで登録内容を確認

- 以下の場合以外は、submit 後にアクセッション番号が自動発行される
- NCBI Taxonomy に登録されていない新規生物名
 - locus_tag_prefix (ゲノム上の遺伝子を一意に識別する locus_tag の先頭文字列) が記載されている

[\(August 14 to 17\) DDBJ Center Closing for Summer Break](#)

[Revision of the Terms of Use \(Access and Benefit-Sharing\)](#)

Account: test09

[登録ナビゲーションサイト](#): データに関するいくつかの質問に答えると登録方法が案内されます。

[Submission navigation site](#): By answering several questions about your data, you will be guided how to submit your data.

[Register a public key](#) to submit data to DRA/GEA.

DRA/GEA にデータを登録するためには、[公開鍵の登録](#)が必要です。

When you are submitting data derived from human subjects, please read "[Submission of research data from human subjects](#)".

ヒトに由来するデータを登録する場合は「[ヒトを対象とした研究データの登録について](#)」を熟読してください。

❗ Use only ASCII characters (English letters) throughout your submission. Non-ASCII characters (for example, Japanese and special characters) are not accepted.

登録には ASCII 文字 (英語文字) のみを使用してください。非 ASCII 文字 (日本語や特殊文字など) は受け付けていません。

❗ If there is no reply from submitters after three months of initial contact from us, submissions will be cancelled.

登録者に問い合わせた後三か月以上回答が無い場合は Submission をキャンセルいたします。

[BioProject](#)

A collection of biological data related to a single initiative, originating from a single organization or from a consortium.

You can register a project here and later submit and link sequence data to the project. [Manual](#)

ここからプロジェクトのみを登録し、後から配列データを登録してプロジェクトに関連付けることができます。 [マニュアル](#)

[BioSample](#)

Descriptions of biological source materials used to generate experimental data in any of DDBJ's primary data archives.

You can register sample(s) here and later submit and link sequence data to the sample(s). [Manual](#)

ここからサンプルのみを登録し、後から配列データを登録してサンプルに関連付けることができます。 [マニュアル](#)

[DDBJ Sequence Read Archive \(DRA\)](#)

DRA stores raw and aligned sequence data from next-generation sequencing platforms.

You can complete both your BioProject as well as BioSample submissions within DRA submission interface. [Manual](#)

DRA 登録インターフェースでまとめて BioProject, BioSample と DRA メタデータを登録することができます。 [マニュアル](#)

DRA 登録: ファイルアップロード

DRA Submission ID : test09-0001

DRA Submission ID	DRA Accession	Status	Created Date	Hold Date
test09-0001		New	2026-08-04	-----

❗ All metadata (Experiment, Run, Analysis and referencing BioProject and BioSample) and sequencing data in this submission will be released at the same time.

この Submission に含まれる全てのメタデータ (Experiment, Run, Analysis, 参照されている BioProject と BioSample) とシーケンスデータは同時に公開されます。

❗ If your submission has more than 100 Experiments or Runs, we recommend to register metadata by uploading XML files.

[Manual](#)

ExperimentやRunが100件を超える登録の場合、XMLファイルをアップロードしてメタデータを登録することを推奨します。[マニュアル](#)

1. Upload data files to ftp-private.ddbj.nig.ac.jp:/submission/test09/test09-0001
データファイルを次のディレクトリにアップロードしてください ftp-private.ddbj.nig.ac.jp:/submission/test09/test09-0001

2. Enter or update metadata in the web interface.
ウェブインターフェースでメタデータを新規入力もしくは更新してください。

Enter / Update metadata

3. Validate uploaded data files
アップロードしたデータファイルを検証処理にかけてください。

Validate data files Stop validation

* Stop validation to edit metadata or upload data files
検証処理を停止してからメタデータを編集もしくはデータファイルをアップロードして

sftp で fastq をアップロード
<https://www.ddbj.nig.ac.jp/upload.html>

```
sftp -i private_key test09@ftpprivate.ddbj.nig.ac.jp

sftp>
sftp> cd test09-0001
sftp> put illumina_1.fq.gz
sftp> put illumina_2.fq.gz
sftp> put pacbio.fq.gz
```



登録順序

DRA 登録: メタデータ作成

1. Upload data files to ftp-private.ddbj.nig.ac.jp:/submission/test09/test09-0001
データファイルを次のディレクトリにアップロードしてください ftp-private.ddbj.nig.ac.jp:/submission/test09/test09-0001

2. Enter or update metadata in the web interface.
ウェブインターフェースでメタデータを新規入力もしくは更新してください。

Enter / Update metadata

3. Validate uploaded data files
アップロードしたデータファイルを検証処理にかけてください。

Validate data files Stop validation

* Stop validation to edit metadata or upload data files
検証処理を停止してからメタデータを編集もしくはデータファイルをアップロードしてください。

Submit or update metadata by directly uploading XML files.

XML Upload [+]

Component [-]

None

Data Files [-]

File Name	Metadata	Uploaded
illumina_1.fq.gz	no	yes
illumina_2.fq.gz	no	yes
pacbio.fq.gz	no	yes

DRA 登録: SUBMISSION

SUBMISSION

BIOPROJECT

BIOSAMPLE

EXPERIMENT

RUN

ANALYSIS (optional)

Submitting organization

*Submitting organization :

National Institute of Genetic

Specify when this submission should be released to the public.

!

All metadata (Experiment, Run, Analysis and referencing BioProject and BioSample) and sequencing data in this submission will be released at the same time.

Data Release

☒ Release on specified date (not viewable until this date or the release of linked data, whichever is first)

Release date (YYYY-MM-DD)

2027-03-01

☐ Release immediately following curation

Submitter

+ Add another submitter

#	*Name	*Email	Delete
1	Hanako Mishima	test09@nig.ac.jp	
	Ichiro Fuji	test10@nig.ac.jp	

Next (BIOPROJECT)

登録者情報と公開予定

DRA 登録: BIOPROJECT を選択

SUBMISSION

BIOPROJECT

BIOSAMPLE

EXPERIMENT

RUN

ANALYSIS (optional)

Select a BioProject for this DRA submission.

	BioProject Accession	BioProject Submission ID	Title
	filter records Reset	filter records Reset	filter records Reset
☒	PRJDB42877	PSUB049081	Genome sequencing project of American Shorthair cat

If you have not registered your project, [register a project](#).

Next (BIOSAMPLE)

登録した PRJDB42877 を選択

DRA 登録: BIOSAMPLE を選択

SUBMISSIONBIOPROJECTBIOSAMPLEEXPERIMENTRUNANALYSIS (optional)

Select sample(s) for this DRA submission.

Select filtered BioSamplesUnselect filtered BioSamples

	BioSample Accession	BioSample Submission ID	Title
	filter records Reset	filter records Reset	filter records Reset
☒	SAMD01949766	SSUB048035 : Felis catus American Shorthair breed	Felis catus American Shorthair breed

If you have not registered your sample, [register sample\(s\)](#).

Next (EXPERIMENT)

登録した SAMD01949766 を選択

DRA 登録: EXPERIMENT

SUBMISSION

BIOPROJECT

BIOSAMPLE

EXPERIMENT

RUN

ANALYSIS (optional)

• For more information, please see the [Experiment metadata](#).
詳細は [Experiment メタデータ](#)を参照してください。

• Click the "Save" button and fix aliases to edit metadata in TSV file.
"Save" をクリックして alias を確定してから TSV で内容を編集してください。

Save

Next (RUN)

Edit Experiment by using tab-delimited text (TSV) file

Download TSV file

Upload TSV file

ファイルを選択

選択されていません

Experiment

Add another Experiment(s)

1

Copy Experiment #1

#	Alias	*BioSample Used	Library Name	*Library Source	*Library Selection
1	New	SAMD01949766	Senzu Illumina	GENOMIC	RANDOM
	New	SAMD01949766	Senzu PacBio	GENOMIC	RANDOM

Save

Next (RUN)

*Library Strategy	Library Construction Protocol	*Instrument	*Library Layout	Insert Size	Delete	Delete All
WGS	DNeasy Blood & Tissue Kit (QIAGEN), TruSeq DNA PCR-F	HiSeq X Five	paired			
WGS	DNeasy Blood & Tissue Kit (QIAGEN), SMRTbell express t	Sequel	single			

1つの BioSample に由来する2つの「ライブラリー + シークエンス機種」

51

DRA 登録: RUN

SUBMISSION

BIOPROJECT

BIOSAMPLE

EXPERIMENT

RUN

ANALYSIS (optional)

• For more information, please see the [Run metadata](#).

詳細は [Run メタデータ](#) を参照してください。

• Click the "Save" button and fix aliases to edit metadata in TSV file.

"Save" をクリックして alias を確定してから TSV で内容を編集してください。

Save

Select data files for Run

Edit Run by using tab-delimited text (TSV) file

Download TSV file

Upload TSV file

ファイルを選択

選択されていません

Run

Add another Run(s)

1

#	Alias	*Experiment Referenced	Delete
1	New	test09-0001_Experiment_0001	
2	New	test09-0001_Experiment_0002	

Save

Select data files for Run

Edit file by using tab-delimited text (TSV) file

Download TSV file

Upload TSV file

ファイルを選択

選択されていません

File

File Type: "generic_fastq": fastq with variable read length, "fastq": fastq with constant read length.

#	Run/Analysis	File Name	*Run/Analysis contains files	*File Type	Copy #1	*MD5 Checksum
1	Run	illumina_1.fq.gz	test09-0001_Run_0001	fastq		680260f7b5e321efb963a5bb5499dffe
2	Run	illumina_2.fq.gz	test09-0001_Run_0001	fastq		a4ed6a5533211cb1531434e79f096d17
3	Run	pacbio.fq.gz	test09-0001_Run_0002	fastq		e39f509b12c1fa538e39d54ff236c897

Save

Back to Run

Run 1: Illumina paired – illumina_1.fq.gz, illumina_2.fq.gz

Run 2: Sequel single – pacbio.fq.gz

MD5 チェックサム <https://www.ddbj.nig.ac.jp/checksum.html>

DRA 登録: メタデータを登録

Submit/Update DRA metadata

☒ I confirm that, to the best of my knowledge, the data submitted here are not subject to any restrictions that would prevent their open sharing through the databases operated by the BioData Science Initiative (BSI). I have also reviewed and agreed to the BSI Terms of Use, including the provisions on the rights and duties of submitters, the handling of human-derived research data, and the access and benefit-sharing (ABS) of genetic resources.

Read more:

- BSI Terms of Use: <https://www.ddbj.nig.ac.jp/policies-e.html>
- Applicable national and international laws and rules on access and benefit-sharing: <https://www.ddbj.nig.ac.jp/faq/en/abs-data-sharing-e.html>

SUBMISSION

BIOPROJECT

BIOSAMPLE

EXPERIMENT

RUN

ANALYSIS (optional)

Data files for Run/Analysis

- Click the "Save" button and fix aliases to edit metadata in TSV file.
In the TSV file, enter "Run" in the leftmost "Run/Analysis" column to distinguish the data files.
"Save" をクリックして alias を確定してから TSV で内容を編集してください。
TSV 左端の "Run/Analysis" カラムにはファイルを区別するため "Run" と入力してください。
- Associate the uploaded data files to Run or Analysis. Use TSV file to sort and associate many files.
アップロードしたファイルを Run/Analysis と関連付けてください。TSV ファイルを使ってデータファイルを並べ替えたうえで関連付けることができます。
- If you are not submitting Analysis data, please submit metadata by clicking the [Submit/Update DRA metadata] at here.
Analysis データを登録しない場合は、ここで [Submit/Update DRA metadata] をクリックしてメタデータを投稿してください。

Save

Back to Run

DRA 登録: データファイルの検証

Submission Validated

DRA Submission ID : test09-0001

DRA Submission ID	DRA Accession	Status	Created Date	Hold Date
test09-0001		Metadata Submitted	2026-08-04	2027-03-01

❗ All metadata (Experiment, Run, Analysis and referencing BioProject and BioSample) and sequencing data in this submission will be released at the same time.
この Submission に含まれる全てのメタデータ (Experiment, Run, Analysis, 参照されている BioProject と BioSample) とシーケンスデータは同時に公開されます。

❗ If your submission has more than 100 Experiments or Runs, we recommend to register metadata by uploading XML files.
[Manual](#)
ExperimentやRunが100件を超える登録の場合、XMLファイルをアップロードしてメタデータを登録することを推奨します。[マニュアル](#)

1. Upload data files to ftp-private.ddbj.nig.ac.jp:/submission/test09/test09-0001
データファイルを次のディレクトリにアップロードしてください ftp-private.ddbj.nig.ac.jp:/submission/test09/test09-0001

2. Enter or update metadata in the web interface.
ウェブインターフェースでメタデータを新規入力もしくは更新してください。

Enter / Update metadata

3. Validate uploaded data files
アップロードしたデータファイルを検証処理にかけてください。

Validate data files

Stop validation

* Stop validation to edit metadata or upload data files
検証処理を停止してからメタデータを編集もしくはデータファイルをアップロードしてください。

Upload and validate run data files along with the submitted metadata to start reviewing process.
This submission will be reviewed sequentially when the status become "submission_validated"
メタデータの投稿とアップロードしたファイルの検証処理が終了した後、データ査定が開始されます。
この Submission はステータスが "submission_validated"になると順次査定されます。

データファイル形式の検証処理をパスすると、ステータスが「Submission validated」になり、メールで通知される

DRA 登録: アクセッション番号発行

DRA Submission ID : test09-0001

DRA Submission ID	DRA Accession	Status	Created Date	Hold Date
test09-0001	DRA029513	Private	2026-08-04	2027-03-01 Change

ⓘ All metadata (Experiment, Run, Analysis and referencing BioProject and BioSample) and sequencing data in this submission will be released at the same time.

この Submission に含まれる全てのメタデータ (Experiment, Run, Analysis, 参照されている BioProject と BioSample) とシーケンスデータは同時に公開されます。

ⓘ If your submission has more than 100 Experiments or Runs, we recommend to register metadata by uploading XML files.

[Manual](#)

ExperimentやRunが100件を超える登録の場合、XMLファイルをアップロードしてメタデータを登録することを推奨します。[マニュアル](#)

⋮

Component [-]				
Object	BioProject ID	BioSample ID	DRA Accession	Alias
submission			DRA029513	test09-0001_Submission
+ experiment	PRJDB42877	SAMD01949766	DRX1055790	test09-0001_Experiment_0001
+ run			DRR1080205	test09-0001_Run_0001
+ experiment	PRJDB42877	SAMD01949766	DRX1055791	test09-0001_Experiment_0002
+ run			DRR1080206	test09-0001_Run_0002

DRR Run アクセッション番号を論文や後の登録で引用

データを登録する

BioProject

BioSample

DRA (DDBJ SRA)

DDBJ

DDBJ: アセンブリ配列の登録窓口

データベース	登録システム	対象	アカウント
BioProject	D-way	プロジェクト情報	DDBJ アカウント
BioSample		サンプル情報	
DRA		NGS リード	
DDBJ	NSSS: ウェブ塩基配列登録システム	配列数100以下、かつ、配列長500kb以下	アカウント不要
	MSS: ファイルベースの登録システム	上記基準以上の大規模配列（ゲノム等）	DDBJ アカウント



- 登録ファイル作成（アノテーションと配列ファイル）
- 登録ファイルの事前チェック
- MSS フォームにログイン
- MSS フォームにデータ種別を記入し、ファイルをアップロード
- Mass-ID（管理用仮 ID）発行
- 査定（メールでの問い合わせ）
- アクセッション番号発行

DDBJ: アノテーションファイルと配列ファイル

Entry	Feature	Location	Qualifier	Value
COMMON	SUBMITTER		ab_name	Fuji,I.
			ab_name	Mishima,H.
			contact	Hanako Mishima
				
			sequence read archive	DRR1080206
	DATE		hold_date	20270301
AnAms1.0_A1	source	1..129629249	chromosome	A1
			isolate	Senzu
			mol_type	genomic DNA
			note	breed: American Shorthair
			organism	Felis catus
			sex	female
	assembly_gap	107532..107631	estimated_length	unknown
			gap_type	within scaffold
			linkage_evidence	proximity ligation
				
AnAms1.0_A2	source	1..174768116	chromosome	A2
				

```
>AnAms1.0_A1
TCCAAGATTCCCAGGGTTGGGGAAGGGTGGCTTCCAGGCTGTTAGCAGCAGAGTTGAGGG
...
...
...
...
...
GAGGGAGAATCCCAAGCAGGCTCCACGCTGCCAGTGCAGAGTCCCACATGGGGT
//
>AnAms1.0_A2
CCCCAAGGACCTGTGCTGGACGCTGAAATGCCCCGCCTGGGACCAGGTGTGCCCTCAGG
```

アノテーションファイル

- メタデータを記載する5列のタブ区切りテキストファイル
- COMMON: 共通情報を記載（登録者、文献、公開予定日等）
- ENTRY: 各配列の情報を記載（Feature/Location/Qualifier）

配列ファイル

- FASTA 形式で記載された塩基配列
- 「>」で始まるヘッダー行
- 配列間は「//」で区切る
- エントリ名でアノテーションファイルとリンク（例 AnAms1.0_A1）

アノテーションファイル: COMMON 共通情報

COMMON	SUBMITTER		ab_name	Fuji,I.
			ab_name	Mishima,H.
			contact	Hanako Mishima
			email	test09@nig.ac.jp
			institute	National Institute of Genetics
			country	Japan
			state	Shizuoka
			city	Mishima
			Street	1111 Yata
			zip	411-8540
	REFERENCE		title	AnAms1.0; A high quality chromosome-scale assembly of an American Shorthair Cat
			ab_name	Fuji,S.
			ab_name	Mishima,H.
			year	2020
			status	Published
	ST_COMMENT		tagset_id	Genome-Assembly-Data
			Assembly Method	FALCON v. 1.8.8; HiRise; Bionano Access v. 1.4
			Assembly Name	AnAms_1.0
			Genome Coverage	120x
			Sequencing Technology	PacBio Sequel; Illumina HiSeqX
	DBLINK		project	PRJDB42877
			biosample	SAMD01949766
			sequence read archive	DRR1080205
			sequence read archive	DRR1080206
	DATE		hold_date	20270301

登録者

文献

配列決定・アセンブリ手法

BioProject
BioSample

DRA Run

公開予定日

アノテーションファイル: ENTRY 配列情報

Entry	Feature	Location	Qualifier	Value
AnAms1.0_A1	source	1..129629249	chromosome	A1
			isolate	Senzu
			mol_type	genomic DNA
			note	breed: American Shorthair
			organism	Felis catus
			sex	female
	assembly_gap	107532..107631	estimated_length	unknown
			gap_type	within scaffold
			linkage_evidence	proximity ligation
			⋮	
			⋮	
AnAms1.0_A2	source	1..174768116	chromosome	A2
			⋮	

生物・サンプル・配列分子

配列アノテーション

次のエントリ（配列）

アノテーションツールの出力結果を加工して作成

配列ファイル: fasta 形式

```
>AnAms1.0_A1
TCCAAGATTCCCAGGGTTGGGGAAGGGTGGCTTCCAGGCTGTTAGCAGCAGAGTTGAGGG
GACATTTTCTTTCAGTCTAGGTCATGTTTTTCTGCTTTCAGTTCACGTCCCCAGCAGAG
GGAATCTCATTTGGCCACCTTGGCCTAAAGTCTCTACCCTTGGAACAACCCAAGCCAGGT
...
GAGGGAGAATCCCAAGCAGGCTCCACGCTGCCAGTGCAGAGTCCCACATGGGGT
//
>AnAms1.0_A2
CCCCCAAGGACCTGTGCTGGACGCTGAAATGCCCCGCCTGGGACCAGGTGTGCCCTCAGG
TGCGCCCGCTCCCTAGCAATGAGAGAGAGAGAGAGAGAGAGAGGGAGTGAGACAAGGAGGA
...
TCATGAAGATAAAAAGCTTCTGCACTGCAAAGGAAACAAGCAACAAAACCTAAAAGG
//
...
```

エントリ名で配列とアノテーションファイル中のメタデータが結びつく

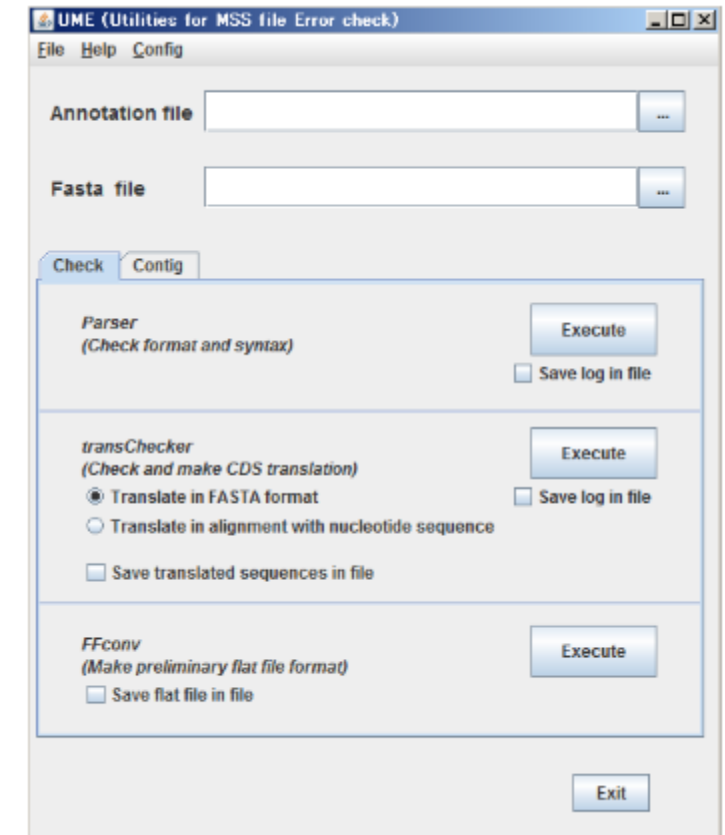
ファイルの事前チェック

UME (Utilities for MSS file Error check)

- Parser: アノテーションファイルと配列ファイルの構文とフォーマットをチェック
- transChecker: CDS (タンパク質コード配列) のアミノ酸翻訳をチェック

チェックツール UME をダウンロード、ファイルを指定してチェックを実行。

エラーのないファイルを送付することが、円滑な登録に重要。



MSS フォーム: 新規登録申請

MSS 登録フォーム

Mass Submission System (MSS) では、[以下 1 - 3 のいずれかに該当する配列データ登録ファイル\(reads ではない\)](#)を受け付けます。
[DDBJ に登録可能なデータ](#)も合わせてお読みください。

1. 種別、または、量が以下に該当する配列データ

◦ [EST](#)、[TSA](#)、[HTC](#)、[GSS](#)、[HTG](#)、[WGS](#)、[TLS](#)

▪ データ種別の詳細は、[Annotated/Assembled データの区分](#)をご確認ください。

◦ 1 配列が長い場合、目安は 500 kb 以上

◦ 1 配列あたりに多数の Feature がある場合、目安は 30 以上

◦ 登録する配列の数が多い場合、総計 100 を超えるとき

2. finished/draft の別を問わずレプリコン全長規模相当の配列データ

3. DBLINK に BioProject, BioSample の記載が必要な配列データ

登録を始めるには [D-way アカウント](#) でログインしてください。

D-way アカウントでログイン

[ユーザ名またはパスワードを忘れてしまった新規アカウント登録](#)

新規登録申請

1. 前提条件の確認

2. 登録ファイルの選択

3. 情報の入力

4. 申請内容の確認

5. 完了

以下のすべてに該当する場合のみ、[一度の submission として申し込みできます](#)。
いずれか該当しない場合、各 submission が条件に合致するように submission ファイルを分け、別々に申し込みを行ってください。[複数回に分けての申し込みが必要な例](#)もご覧ください。

• コンタクトパーソンが同一である

• 登録データ種別が1種類である

• 公開予定日が同一である

今回登録する配列について当てはまるものを選択してください。

☒ 自身の研究で配列決定した塩基配列である

☐ 公開されているデータからアセンブルした塩基配列 (TPA) である [TPA とは](#)

次へ

MSS フォーム (<https://mss.ddbj.nig.ac.jp>)
DDBJ アカウントでログイン

「自身の研究で配列決定した塩基配列である」を選択

63

MSS フォーム: アップロード

新規登録申請

1. 前提条件の確認

2. 登録ファイルの選択

3. 情報の入力

4. 申請内容の確認

5. 完了

Submission ファイルの送付方法を選択してください。

登録ファイルをアップロードする場合は、事前に[UME \(Parser, transChecker\)](#)にてフォーマットやCDSフィーチャーの翻訳チェックを行い、エラーのないファイルを送付してください。公開予定日を指定するには、[アノテーションファイルCOMMON領域にhold_dateの記載が必要です](#)。フォームへの入力で指定することはできません。

☐ DFASTのジョブIDから登録ファイルを取り込んで送付する

☒ 登録ファイルをこのMSSフォームからアップロードする

☐ ファイル受付用sftpサーバ /[submission/test09/mass](#) にアップロード済みの全登録ファイルを送付する

- massディレクトリに登録ファイルのみを設置後、このオプションを選択してください
- ファイル数やサイズが大きい場合、ロード開始画面が現れるまで数分かかります

[① データのアップロード](#)

アノテーションファイルと塩基配列ファイルでペアになるようにアップロードしてください。複数のペアをアップロード可能です（例えば10個の微生物完全ゲノム配列を登録する場合には、合計20個のファイルを1回のアップロードで登録申請できます）。ファイル名から拡張子を除いた名称が同一なファイルがペアになります。読み込めるファイル名拡張子は次の通りです。圧縮ファイルをアップロードしないでください。

アノテーションファイル

.ann .annt.tsv .ann.txt

塩基配列ファイル

.fasta .fsa .seq.fa .fa .fna .seq

ファイルを選択

するか、ここにファイルをドロップしてください。

✓ AnAms1.0.ann 76.02 kB

コンタクトパーソン Hanako Mishima <test09@nig.ac.jp> (National Institute of Genetics)

公開予定日 2027-03-01

⊖

✓ AnAms1.0.fasta 2.53 GB

エントリ数 20

⊖

ファイルの追加

2 ファイル、2.53 GB

戻る

次へ

Annotation と fasta ファイルをアップロード

64

MSS フォーム: データ種別の入力

新規登録申請

1. 前提条件の確認

2. 登録ファイルの選択

3. 情報の入力

4. 申請内容の確認

5. 完了

合計エントリ数と公開予定日、コンタクトパーソンは、submission fileから取得して表示しています。合致しない場合は、「2. 登録ファイルの選択」に戻り、修正したファイルで転送をやり直してください。

すべての配列ファイルの合計エントリ数と公開予定日 (Hold date)

エントリ数

公開予定日

20

2027/03/01

コンタクトパーソン

E-mail アドレス

フルネーム

test09@nig.ac.jp

Hanako Mishima

所属名 (英語表記)

National Institute of Genetics

塩基配列決定に使用したシーケンサーのタイプ 必須

☒ NGS

☐ Sanger dideoxy sequencing

データ種別 必須

GNM: Finished Level Genome sequence, non-WGS

▼

データ種別を選択してください。種別を判断できない場合、登録ナビゲーションツールをご利用ください。

通知メールの言語 必須

☒ 日本語

☐ 英語

戻る

次へ

シーケンサーのタイプ

データ種別

MSS フォーム: 申請内容の確認

新規登録申請

- 1. 前提条件の確認
- 2. 登録ファイルの選択
- 3. 情報の入力
- 4. 申請内容の確認**
- 5. 完了

塩基配列

- 自身の研究で配列決定した塩基配列である
- 登録ファイルをこのMSSフォームからアップロードする

登録ファイル

AnAms1.0.ann

AnAms1.0.fasta

エントリ数

20

通知メールの言語

日本語

☒ 私は、ここで登録するデータが、私の知る限り、バイオデータ研究拠点（BSI）が運用するデータベースを通じて制限なく公開・共有できることを承認します。また、登録者の権利と義務、ヒト由来研究データの取り扱い、ならびに遺伝資源のアクセスと利益配分（ABS）に関する規定を含むBSI利用規約を確認し、これに同意します。

詳しくはこちら:

- [BSI利用規約](#)
- [アクセスと利益配分に関する国内外の法令・ルール](#)

入力内容をご確認のうえ、よろしければ登録申請ボタンを押してください。クリック後は修正できません。

戻る

登録申請



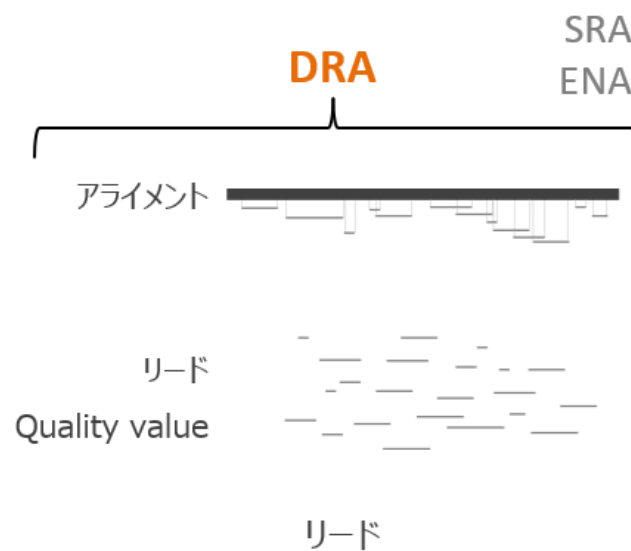
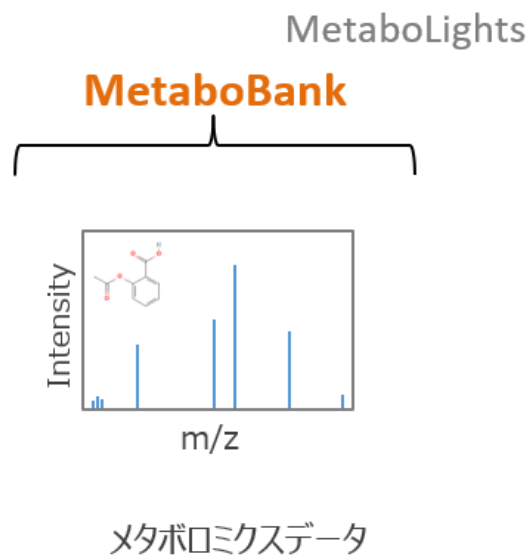
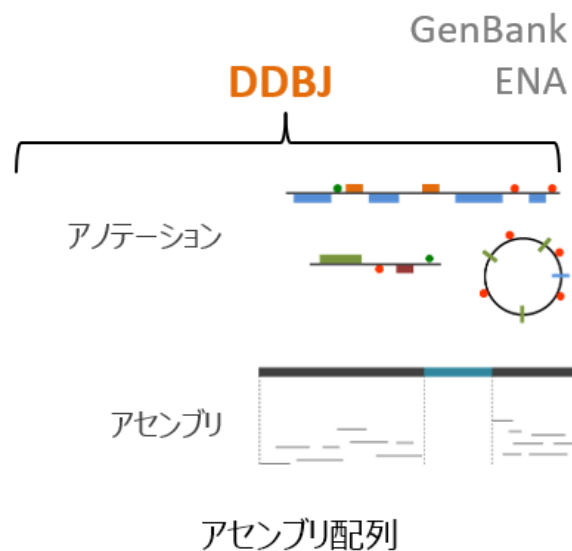
申し込みを受け付けました。MSS申し込みIDは **NSUB000111** です。
mass@ddbj.nig.ac.jp より、入力いただいたアドレスとD-wayアカウントに登録されたメールアドレスにメールが通知されます。メール添付でのsubmissionファイル追加送付はご遠慮ください。

DDBJ: アクセション番号

- 申請後、キュレータにより査定が開始される
- 必要に応じてメールでの問い合わせと対応
- 完成したデータに対して各配列にアクセション番号が発行される

Entry name	Accession
AnAms1.0_A1	AP023152
AnAms1.0_A2	AP023153
AnAms1.0_A3	AP023154
AnAms1.0_B1	AP023155
AnAms1.0_B2	AP023156
AnAms1.0_B3	AP023157
AnAms1.0_B4	AP023158
AnAms1.0_C1	AP023159
AnAms1.0_C2	AP023160
AnAms1.0_D1	AP023161
AnAms1.0_D2	AP023162
AnAms1.0_D3	AP023163
AnAms1.0_D4	AP023164
AnAms1.0_E1	AP023165
AnAms1.0_E2	AP023166
AnAms1.0_E3	AP023167
AnAms1.0_F1	AP023168
AnAms1.0_F2	AP023169
AnAms1.0_X	AP023170
AnAms1.0_unplaced	AP023171

オープンアクセスデータベース



アクセス制限データベース

JGA

Individual-level genotype and phenotype

DBCLS

Data Access Committee (DAC)

BioProject
プロジェクト情報

BioSample
サンプル情報



アクセッション番号の引用

投稿する論文の規定に従ってください。

基本的には登録データに対するアクセッション番号、例えば DDBJ 塩基配列や DRA Run アクセッション番号、の引用を推奨します。

データ全体を言及する場合は BioProject アクセッション番号を引用します。ただし、論文中のデータと登録データが対応付けられるようにメタデータが記載されていることが必要です。

FAQ: 投稿論文ではどのアクセッション番号を引用するべきでしょうか？

<https://www.ddbj.nig.ac.jp/faq/ja/cite-accession.html>

ヒトデータの登録

人を対象とする生命科学・医学系研究に関する倫理指針が適用される研究で得られたヒトデータを登録する場合、NBDC ヒトデータベースへの提供申請が必要になります。

- アクセス制限が必要: JGA に登録
- アクセス制限が不要: BioProject/BioSample/DRA/GEA/MetaboBank に登録



<https://humandbs.dbcls.jp/data-submission>

FAIR



Findable



Accessible



Interoperable



Reusable

Source: National Library of Medicine (NLM), NIH
<https://www.nlm.nih.gov/oet/ed/cde/tutorial/02-200.html>

余裕をもった登録をお願いいたします