

次世代シークエンスデータベース の紹介と歩き方

児玉 悠一

Kodama Yuichi, PhD

国立遺伝学研究所・DDBJ センター、コーディネータ
National Institute of Genetics, DDBJ Center, Coordinator

自己紹介



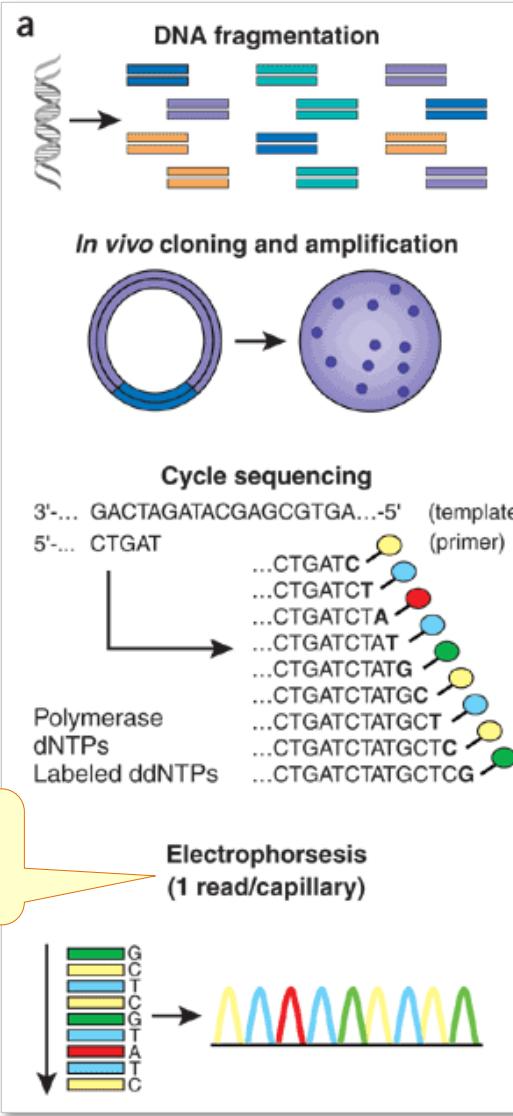
✓ 国立遺伝学研究所 DDBJ センター コーディネータ

講義の内容

1. 次世代シークエンスデータベースと関連データベース
2. Sequence Read Archive (SRA)
3. EBI/NCBI/DDBJ 各データベースの使い方と特徴

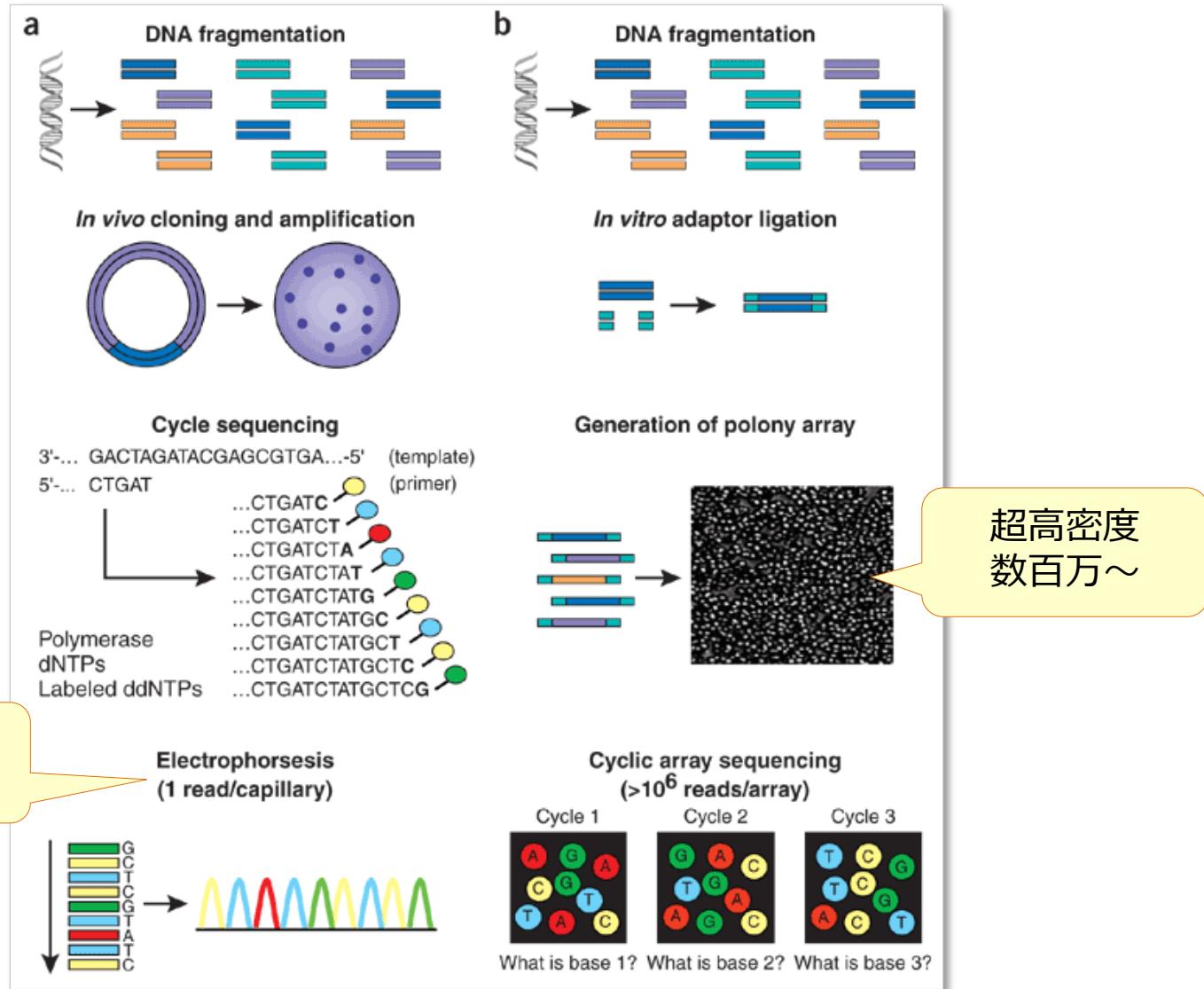
1. 次世代シークエンスデータベースと 関連データベース

サンガー法 v.s. 次世代シークエンス (NGS)

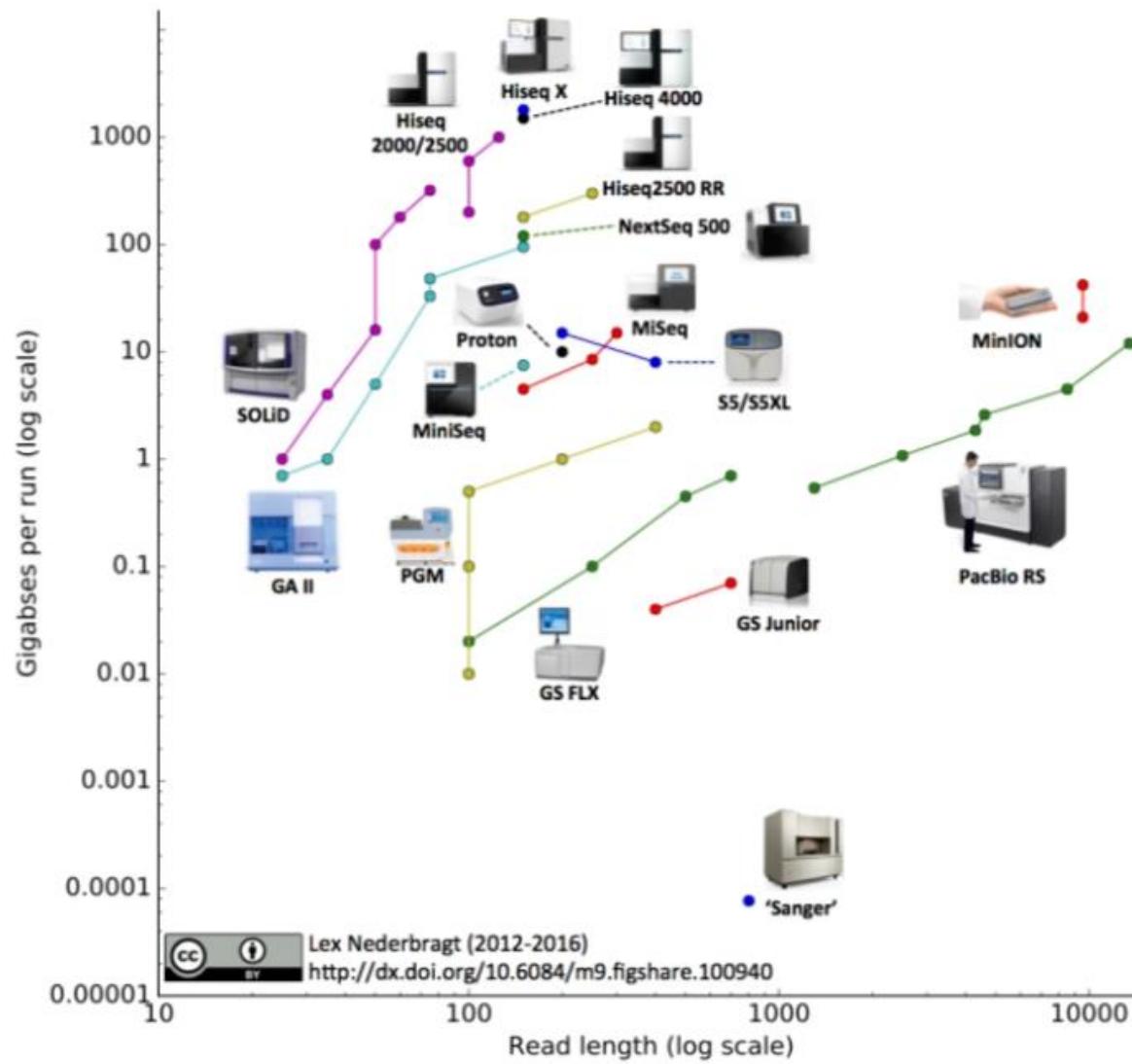


並列解析数
384

サンガーフラッシュ v.s. 次世代シーケンス (NGS)

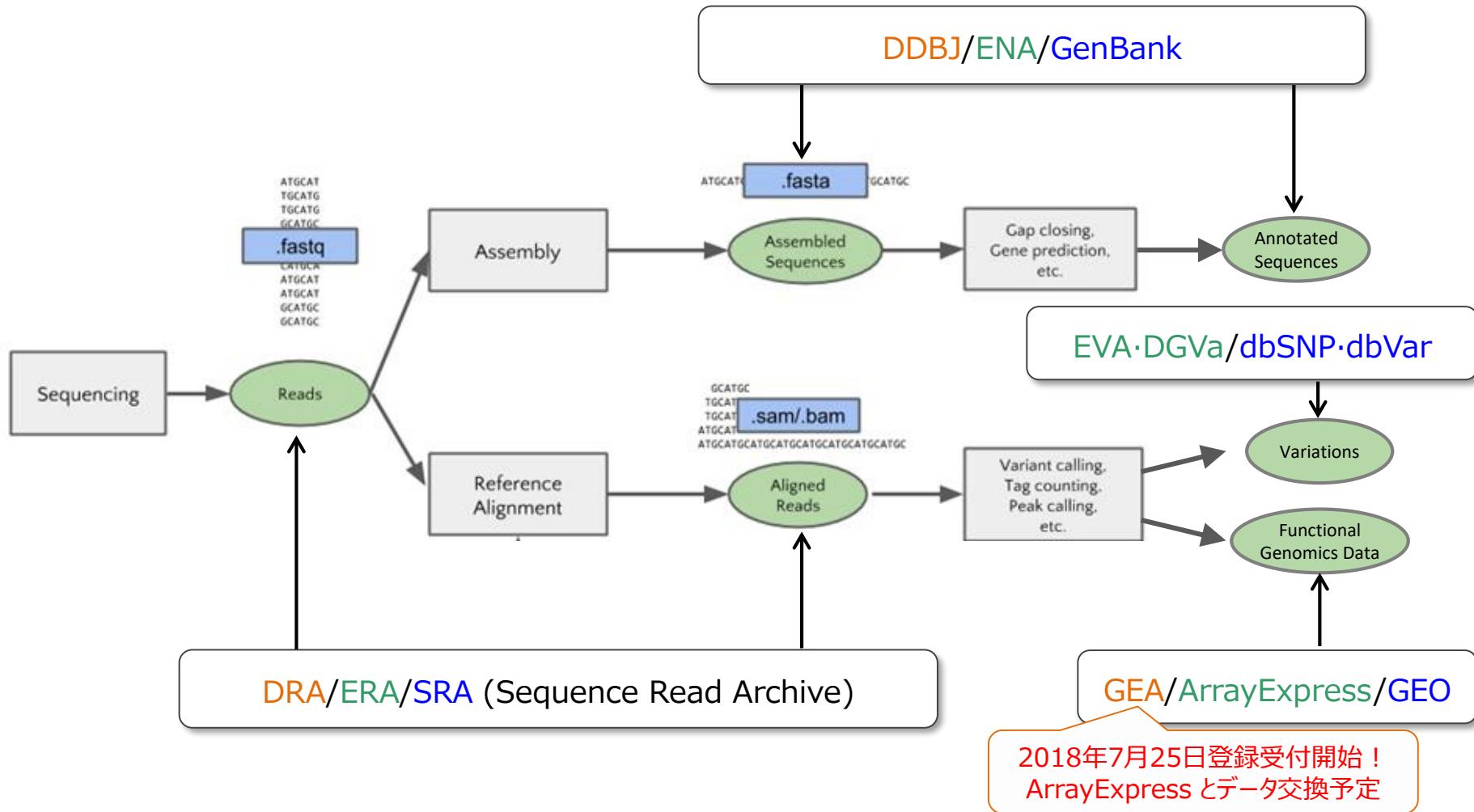


次世代シークエンサー

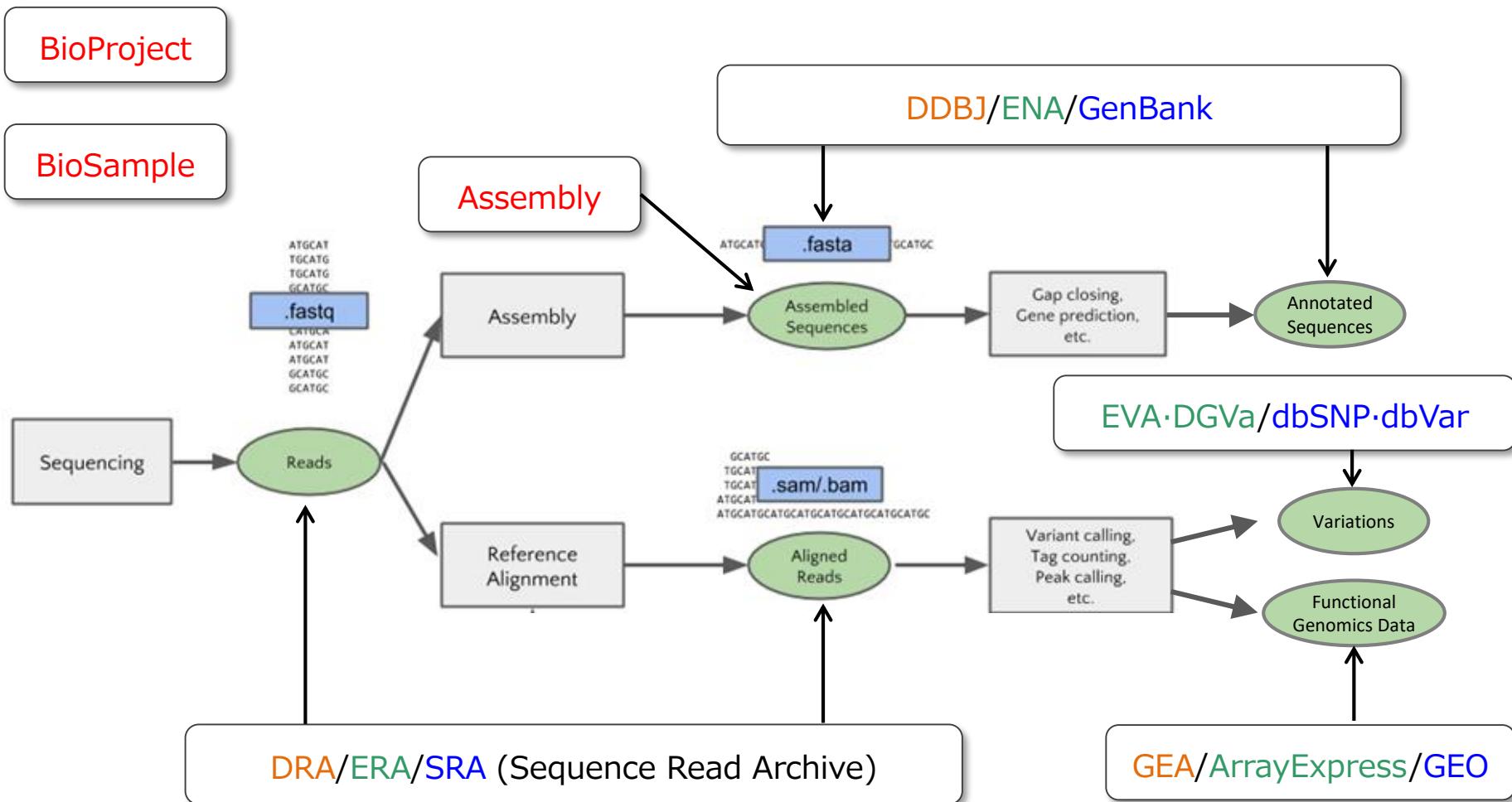


<https://flxlexblog.wordpress.com/2016/07/08/developments-in-high-throughput-sequencing-july-2016-edition/>

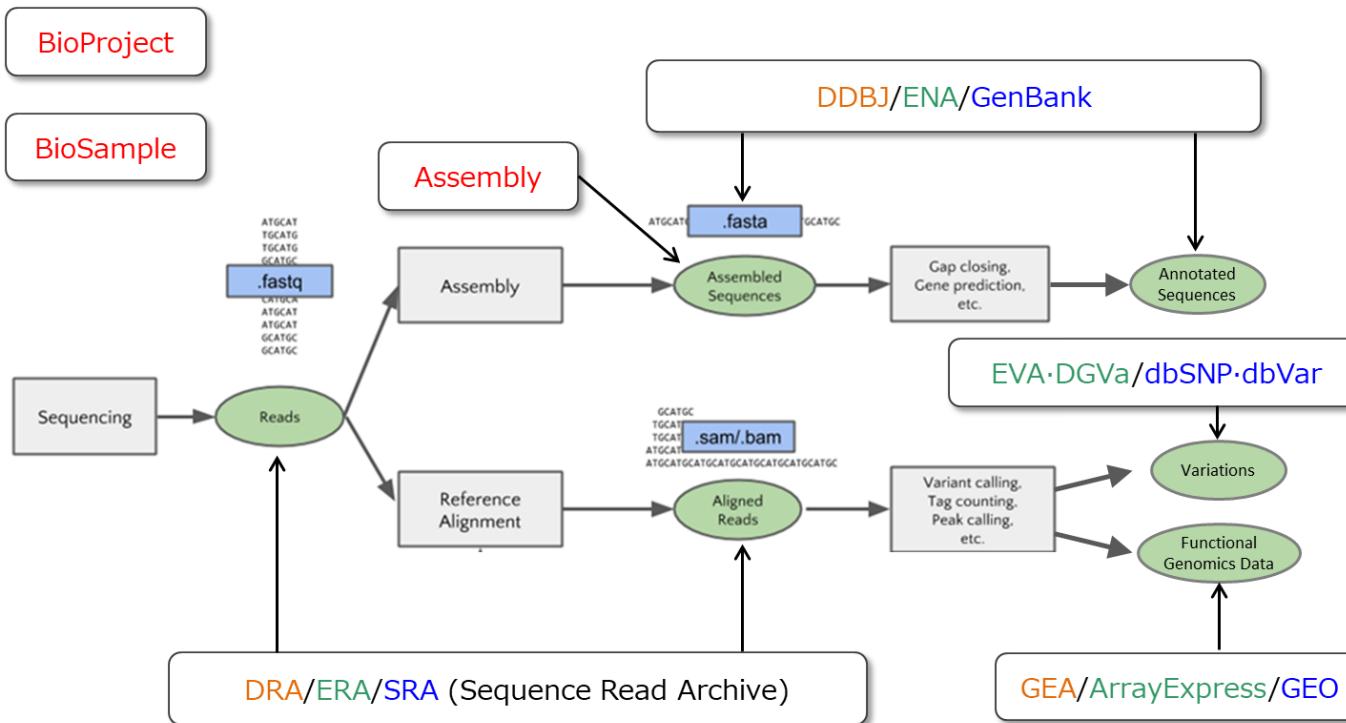
データの種類とデータベース



データを整理するデータベース



アクセス制限データベース



オープンアクセス

アクセス制限



DDBJ Center/EBI/NCBI

INSDC

一次と二次データベース

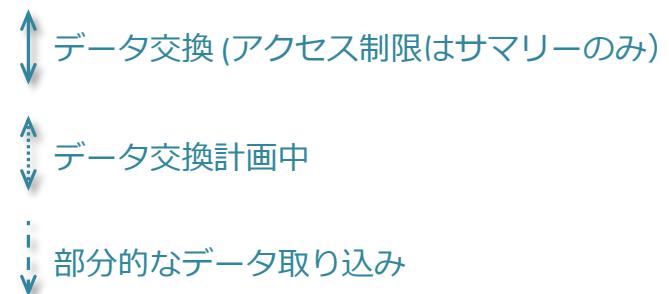
	一次データベース Primary database	二次データベース Secondary database
別の呼び方	Archival database	Curated database; Knowledgebase
データソース	研究者（登録者）が実験で得たデータを直接登録	一次データベースのデータや文献を解析、解釈、キュレーションした結果
例	<ul style="list-style-type: none">• DDBJ/ENA/GenBank• GEA/ArrayExpress/GEO• DRA/ERA/SRA• EVA·DGVa/dbSNP·dbVar• PDB	<ul style="list-style-type: none">• RefSeq• Ensembl• Expression Atlas• ChIP-Atlas• UniProt

EBI Training 資料をもとに作成
<https://www.ebi.ac.uk/training/online/course/bioinformatics-terrified-2018/primary-and-secondary-databases>

一次データベース構築における国際協力

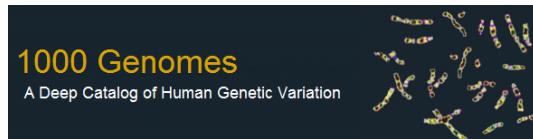
	Annotated sequences	Capillary reads	NGS reads	Study	Sample	Assembly	Functional genomics	Variation	Genotype and phenotype
NCBI	GenBank	Trace Archive	Sequence Read Archive	BioProject	BioSample	Assembly	GEO	dbSNP/dbVar	dbGaP
EBI	European Nucleotide Archive (ENA)						ArrayExpress	EVA/DGVa	EGA
DDBJ	DDBJ	Trace Archive	Sequence Read Archive	BioProject	BioSample	Assembly 準備中	GEA 2018年7月開始		JGA

↔ INSDC (国際塩基配列データベース共同事業)

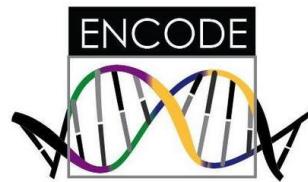


大規模プロジェクトデータベース

オープンアクセス



National Human Genome Research Institute



アクセス制限



THE CANCER GENOME ATLAS

National Cancer Institute

National Human Genome Research Institute



International
Cancer Genome
Consortium



IHEC

International Human Epigenome Consortium



GTEX

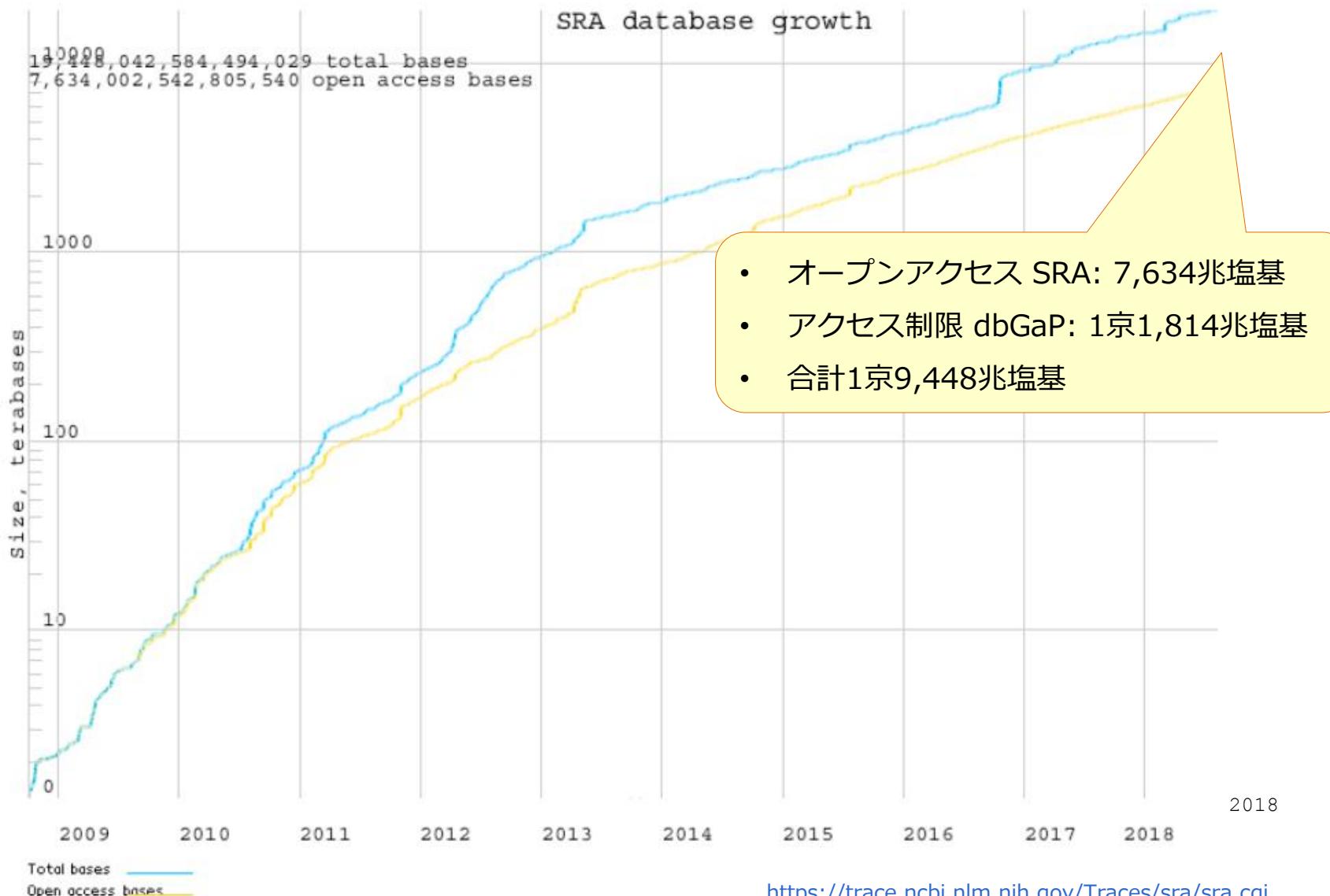


NIH HUMAN
MICROBIOME
PROJECT

- ✓ データの種類毎に一次データベースがある
 - NGS 生・アライメント (SRA)
 - アセンブル・アノテーションされた配列 (DDBJ/ENA/GenBank)
 - 機能ゲノミクスデータ (GEA/ArrayExpress/GEO)
 - 変異 (EVA·DGVa/dbSNP·dbVar)
 - アクセス制限データベース (JGA/EGA/dbGaP)
- ✓ 共通情報を括りだしたデータベース (BioProject/BioSample)
- ✓ 大規模プロジェクトデータベース (ENCODE, TCGA etc)

2. Sequence Read Archive (SRA)

どのくらいのデータがあるか？



大部分は Illumina

SRA Advanced Search Builder

Use the builder below to create your search

[Edit](#)

[Clear](#)

Builder

Platform ▼

[Hide index list](#)

- abi solid (25132)
- bgiseq (1488)
- capillary (2858)
- complete genomics (4765)
- helicos (3989)
- illumina (4164495)**
- ion torrent (60925)
- ls454 (282566)
- oxford nanopore (1912)
- pacbio smrt (37812)

[Previous 200](#)

[Next 200](#)

[Refresh index](#)

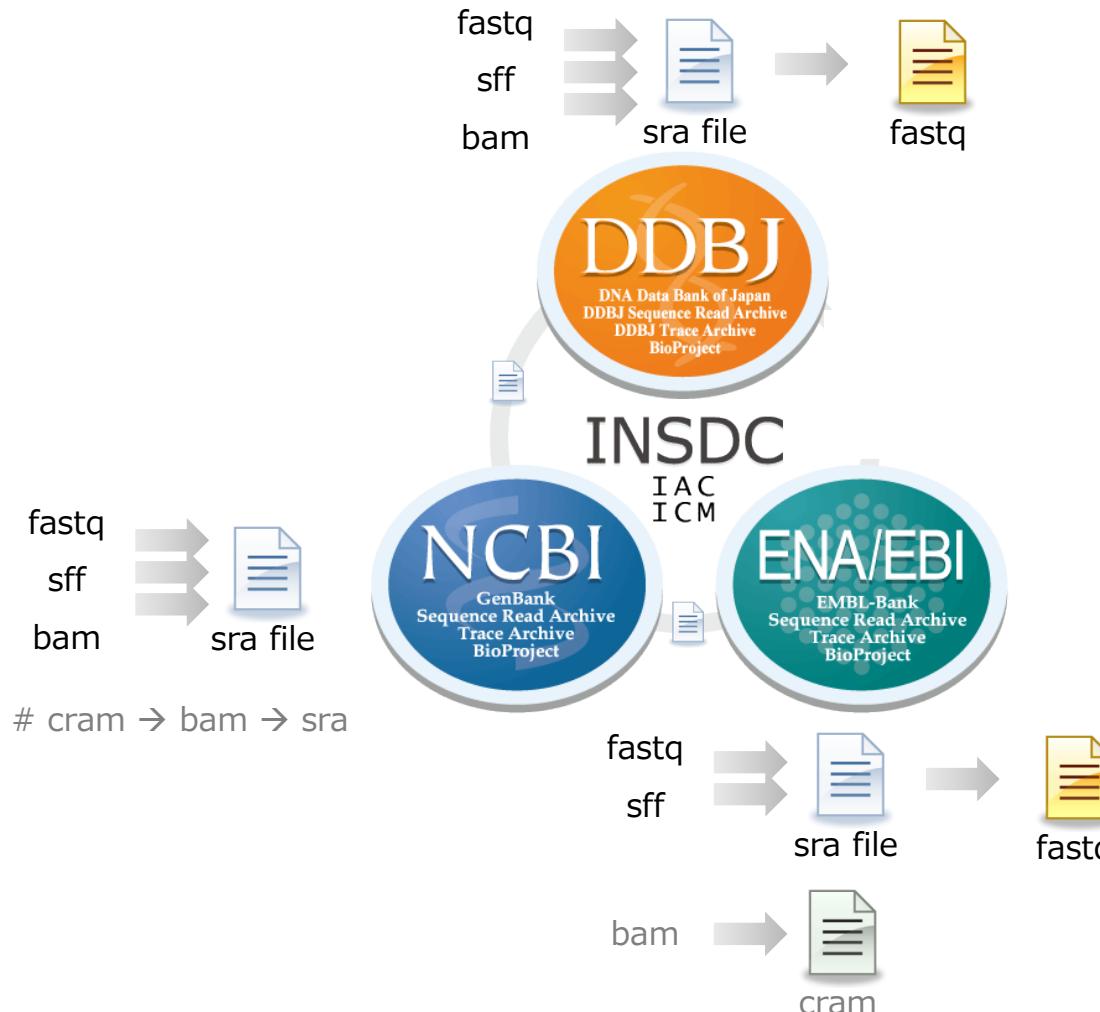
AND ▼ All Fields ▼

[Show index list](#)

[Search](#) or [Add to history](#)

<https://www.ncbi.nlm.nih.gov/sra/advanced>

データファイルフォーマット



- ✓ NCBI が大量リードの格納用フォーマットである sra ファイルとツール一式を開発
- ✓ 三極は sra ファイルでデータ交換
- ✓ NCBI: sra ファイルのみ公開、DDBJ/EBI: fastq も提供

sra ファイルフォーマット

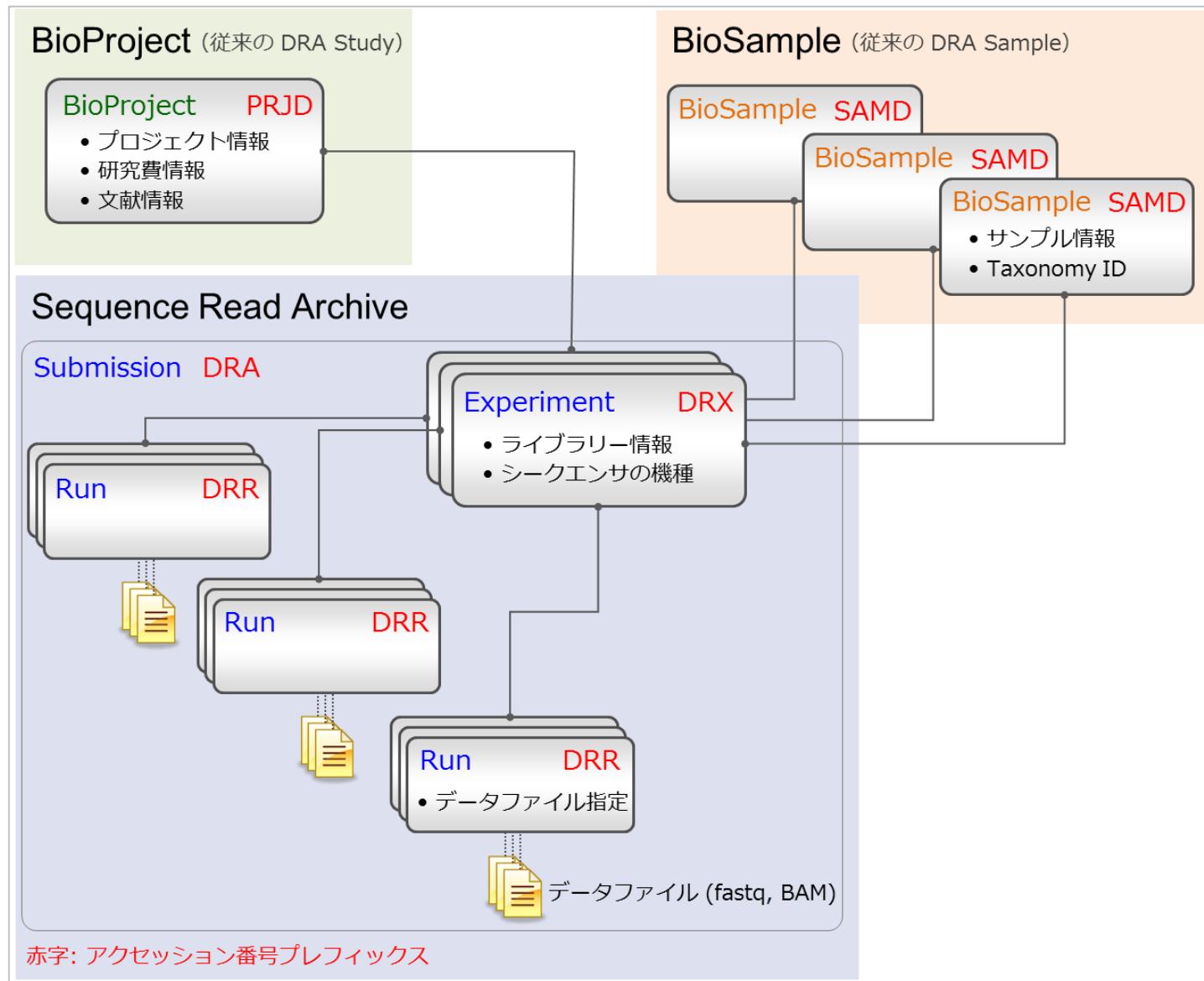
File model	Archive ready?	Stream-able on load?	Auxiliary data?	Run meta data?	Compressed?	Indexed?	Read names parseable ?	Read names indexable?
SRA	Y	Y	Y	Y	Y	Y	Y	Y
BAM	Y	Y	Y	Y	Y	Y	N	N
SFF	Y	Y	Y	Y	Y	Y	Y	Y
HDF5	Y	Y	Y	Y	Y	Y	Y	Y
SOLID	Y	Y	Y	Y	N	N	Y	Y
FASTQ	Y	Y	N	N	N	N	N	N
SRF	N	Y	Y	Y	Y	Y	Y	Y
Illumina native	Y	Y	N	N	N	N	N	N

<https://www.ncbi.nlm.nih.gov/books/NBK242622/>

<https://github.com/ncbi/sra-tools>

- ✓ NCBI が単独で開発
- ✓ sra-tools で fastq, bam 等に変換できる
- ✓ reference genome にアライメントされた bam 由来 sra ファイルは「chromosome 1 のリードだけ取得」といったことができる
- ✓ NCBI は viewer や BLAST で直接 sra ファイルを読み込んでいる

SRA データモデル



プロジェクト概要 →

Ptychodera flava

Whole Genome Shotgun Sequencing of an Hawaiian Acornworm, Ptychodera flava.

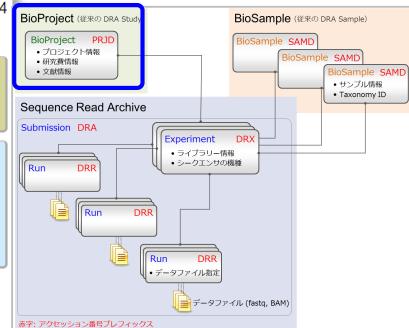
Whole Genome Shotgun Sequencing of an Hawaiian Acornworm, Ptychodera flava. All original genomic library was produced from sperm spawned from one male individual which was sampled near Oahu island in Hawaii at 2006 Dec.
[More...](#)

Accession: PRJDB3182 ID: 302624

See Genome Information for Ptychodera flava

NAVIGATE ACROSS

3 additional projects are related by organism.



生物情報 →

論文情報 →

Accession	PRJDB3182
Data Type	Genome sequencing and assembly
Scope	Monoisolate
Organism	Ptychodera flava [Taxonomy ID: 63121] Eukaryota; Metazoa; Hemichordata; Enteropneusta; Ptychoderidae; Ptychodera; Ptychodera flava
Publications	1. Simakov O <i>et al.</i> , "Hemichordate genomes and deuterostome origins.", <i>Nature</i> , 2015 Nov 18;527(7579):459-65 2. Published online, DOI: 10.1038/nature16150
Submission	Registration date: 24-Nov-2015 Okinawa Institute of Science and Technology
Related Resources	<ul style="list-style-type: none"> Marine Genomics Unit, OIST Marine Biological Laboratory (MBL), Graduate School of Science, Hiroshima University DRA002855 BCFJ01000001-BCFJ01317432
Relevance	Evolution

Project Data:

Resource Name	Number of Links
SEQUENCE DATA	
Nucleotide (total)	218256
WGS master	1
SRA Experiments	11
PUBLICATIONS	
PubMed	1
PMC	1
OTHER DATASETS	
BioSample	1
Assembly	1

Assembly details:

Assembly	Level	WGS	BioSample	Taxonomy
GCA_001465055.1	Scaffold	BCFJ00000000	SAMD00023482	Ptychodera flava

SRA Data Details

Parameter	Value
Data volume, Gbases	287
Data volume, Tbytes	0.23

SRA データサイズ →

NCBI BioProject: <https://www.ncbi.nlm.nih.gov/bioproject/PRJDB3182>

BioProject の役割

プロジェクト概要 →

Ptychodera flava
Whole Genome Shotgun Sequencing of an Hawaiian Acornworm, Ptychodera flava.

Accession: PRJDB3182 ID: 302624

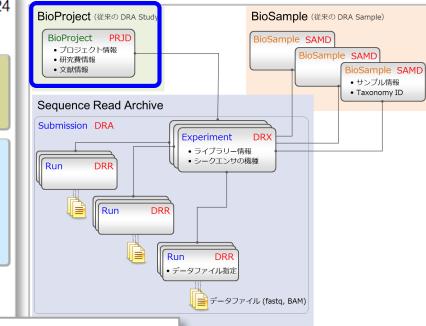
Whole Genome Shotgun Sequencing of an Hawaiian Acornworm, Ptychodera flava. All original genomic library was produced from sperm spawned from one male individual which was sampled near Oahu island in Hawaii at 2006 Dec.
[More...](#)

Accession	PRJDB3182
Data Type	Genome sequencing and assembly
Scope	Monoisolate
Organism	Ptychodera flava [Taxonomy ID: 63121] Eukaryota; Metazoa; Hemichordata; Enteropneusta; Ptychoderidae; Ptychodera; Ptychodera flava

See Genome Information for Ptychodera flava

NAVIGATE ACROSS
3 additional projects are related by organism.

生物情報 →



詳

- ✓ プロジェクトに関する情報
- ✓ 関連データをトップレベルで整理・ナビゲーションの入口
- ✓ 大規模プロジェクトはアンブレラで階層構造になっていることも
- ✓ アクセッション番号プレフィックス

拡

PRJD (DDBJ)/PRJE (EBI)/PRJN (NCBI)

SRA テーパー →

SRA Experiments

PUBLICATIONS	11
PubMed	1
PMC	1
OTHER DATASETS	
BioSample	1
Assembly	1

Assembly details:

Assembly	Level	WGS	BioSample	Taxonomy
GCA_001465055.1	Scaffold	BCFJ00000000	SAMD00023482	Ptychodera flava

SRA Data Details

Parameter	Value
Data volume, Gbases	287
Data volume, Tbytes	0.23

SRA データサイズ →

NCBI BioProject: <https://www.ncbi.nlm.nih.gov/bioproject/PRJDB3182>

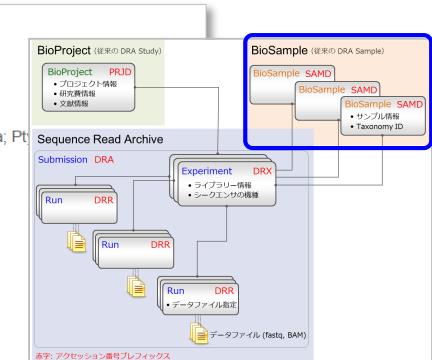
タイトル → **Ptychodera flava Genomic DNAs**

生物情報 → **Organism**
Ptychodera flava
cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Hemichordata; Enteropneusta; Pt;

パッケージ → **Package**
MIGS: eukaryote; version 4.0

サンプル属性 →

Identifiers	BioSample: SAMD00023482																																				
Organism	Ptychodera flava cellular organisms, Eukaryota; Opisthokonta; Metazoa, Eumetazoa; Bilateria; Deuterostomia; Hemichordata; Enteropneusta; Pt;																																				
Attributes	<table border="1"> <tbody> <tr> <td>sample name</td> <td>Ptychodera flava Genomic DNAs</td> </tr> <tr> <td>collection date</td> <td>2006-12-10</td> </tr> <tr> <td>environment biome</td> <td>sea</td> </tr> <tr> <td>environment feature</td> <td>sand</td> </tr> <tr> <td>environment material</td> <td>sea water</td> </tr> <tr> <td>geographic location</td> <td>USA: HI, Oahu</td> </tr> <tr> <td>latitude and longitude</td> <td>NA</td> </tr> <tr> <td>project name</td> <td>Ptychodera flava Genomic DNAs</td> </tr> <tr> <td>isolation and growth condition</td> <td>NA</td> </tr> <tr> <td>reference for biomaterial</td> <td>NA</td> </tr> <tr> <td>number of replicons</td> <td>NA</td> </tr> <tr> <td>estimated size</td> <td>800,000,000</td> </tr> <tr> <td>ploidy</td> <td>haploid</td> </tr> <tr> <td>propagation</td> <td>NA</td> </tr> <tr> <td>cultivar</td> <td>missing</td> </tr> <tr> <td>ecotype</td> <td>missing</td> </tr> <tr> <td>isolate</td> <td>missing</td> </tr> <tr> <td>strain</td> <td>missing</td> </tr> </tbody> </table>	sample name	Ptychodera flava Genomic DNAs	collection date	2006-12-10	environment biome	sea	environment feature	sand	environment material	sea water	geographic location	USA: HI, Oahu	latitude and longitude	NA	project name	Ptychodera flava Genomic DNAs	isolation and growth condition	NA	reference for biomaterial	NA	number of replicons	NA	estimated size	800,000,000	ploidy	haploid	propagation	NA	cultivar	missing	ecotype	missing	isolate	missing	strain	missing
sample name	Ptychodera flava Genomic DNAs																																				
collection date	2006-12-10																																				
environment biome	sea																																				
environment feature	sand																																				
environment material	sea water																																				
geographic location	USA: HI, Oahu																																				
latitude and longitude	NA																																				
project name	Ptychodera flava Genomic DNAs																																				
isolation and growth condition	NA																																				
reference for biomaterial	NA																																				
number of replicons	NA																																				
estimated size	800,000,000																																				
ploidy	haploid																																				
propagation	NA																																				
cultivar	missing																																				
ecotype	missing																																				
isolate	missing																																				
strain	missing																																				
Description	Hawaiian Acornworm, Ptychodera flava, Genomic DNA																																				
	Keywords: GSC:MIxS;MIGS:4.0																																				
Links	marinegenomicsdb																																				
BioProject	PRJDB3182 Ptychodera flava Retrieve all samples from this project																																				
Submission	Okinawa Institute of Science and Technology Marine Genomics Unit ; 2015-11-17																																				
関連データ	Accession: SAMD00023482 ID: 4272872 BioProject SRA Nucleotide																																				



NCBI BioSample: <https://www.ncbi.nlm.nih.gov/biosample/SAMD00023482>

BioSample の役割

タイトル → Ptychodera flava Genomic DNAs

Identifiers BioSample: SAMD00023482

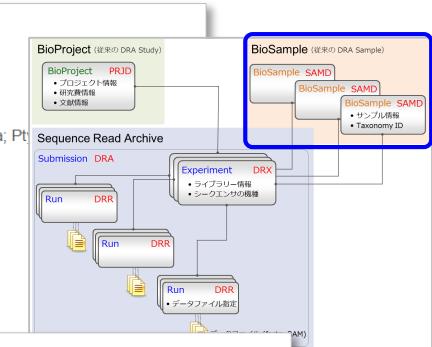
生物情報 → Organism Ptychodera flava

cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Hemichordata; Enteropneusta; Pt

パッケージ → Package MIGS: eukaryote; version 4.0

Attributes

sample name	Ptychodera flava Genomic DNAs
collection date	2006-12-10
environment biome	sea
environment feature	sand



- ✓ データを取得したサンプルが「属性:値」のペアで記述されている
- ✓ サンプルの種類毎のパッケージ（必須・任意属性セット）
- サンプル履歴
 - ✓ サンプルが得られたデータを整理（開発途上）
 - ✓ アクセッション番号プレフィックス

SAMD (DDBJ)/SAME (EBI)/SAMN (NCBI)

Description Hawaiian Acornworm, Ptychodera flava, Genomic DNA

Keywords: GSC:MIxS;MIGS:4.0

Links [marinegenomicsdb](#)

BioProject [PRJDB3182](#) Ptychodera flava
Retrieve [all samples](#) from this project

Submission [Okinawa Institute of Science and Technology Marine Genomics Unit](#); 2015-11-17

Accession: SAMD00023482 ID: 4272872

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

関連データ →

NCBI BioSample: <https://www.ncbi.nlm.nih.gov/biosample/SAMD00023482>

SRA Experiment

DRX025519: Illumina MiSeq paired end sequencing of SAMD00023482

3 ILLUMINA (Illumina MiSeq) runs: 27.5M spots, 14.1G bases, 8Gb downloads

BioProject →

Submitted by: OIST

Study: Whole Genome Shotgun Sequencing of an Hawaiian Acornworm, *Ptychodera flava*

[PRJDB3182](#) • [DRP002816](#) • [All experiments](#) • [All runs](#)

[show Abstract](#)

BioSample →

Sample:

[SAMD00023482](#) • [DRS021946](#) • [All experiments](#) • [All runs](#)

Organism: *Ptychodera flava*

ライブラリー →

Library:

Name: illumina_mp_miseq_5k

Instrument: Illumina MiSeq

Strategy: WGS

Source: GENOMIC

Selection: size fractionation

Layout: PAIRED

Construction protocol:

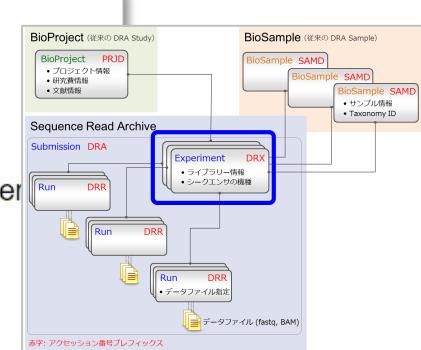
Spot descriptor:



Runs: 3 runs, 27.5M spots, 14.1G bases, [8Gb](#)

Run {

Run	# of Spots	# of Bases	Size	Published
DRR027941	5,307,079	3G	1.4Gb	2015-11-17
DRR027942	16,304,856	7.7G	4.7Gb	2015-11-17
DRR027940	5,904,000	3.4G	1.9Gb	2015-11-17



NCBI Experiment: <https://www.ncbi.nlm.nih.gov/sra/DRX025519>

SRA Experiment の役割

DRX025519: Illumina MiSeq paired end sequencing of SAMD00023482

3 ILLUMINA (Illumina MiSeq) runs: 27.5M spots, 14.1G bases, 8Gb downloads

Submitted by: OIST

Study: Whole Genome Shotgun Sequencing of an Hawaiian Acornworm, *Ptychoderes*

[PRJDB3182](#) • [DRP002816](#) • [All experiments](#) • [All runs](#)

[show Abstract](#)

BioProject →

BioSample →

ライブラリー →

シークエンサー →

- ✓ ライブラリー + シークエンサー

- ✓ BioProject, BioSample と Run をつなぐ中心オブジェクト

- ✓ アクセッション番号プレフィックス

DRX (DDBJ)/ERX (EBI)/SRX (NCBI)

Construction protocol:

Spot descriptor:

1

forward

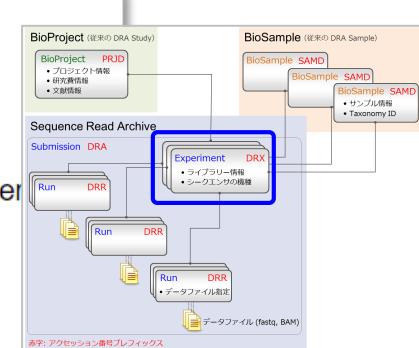
271

reverse

Runs: 3 runs, 27.5M spots, 14.1G bases, [8Gb](#)

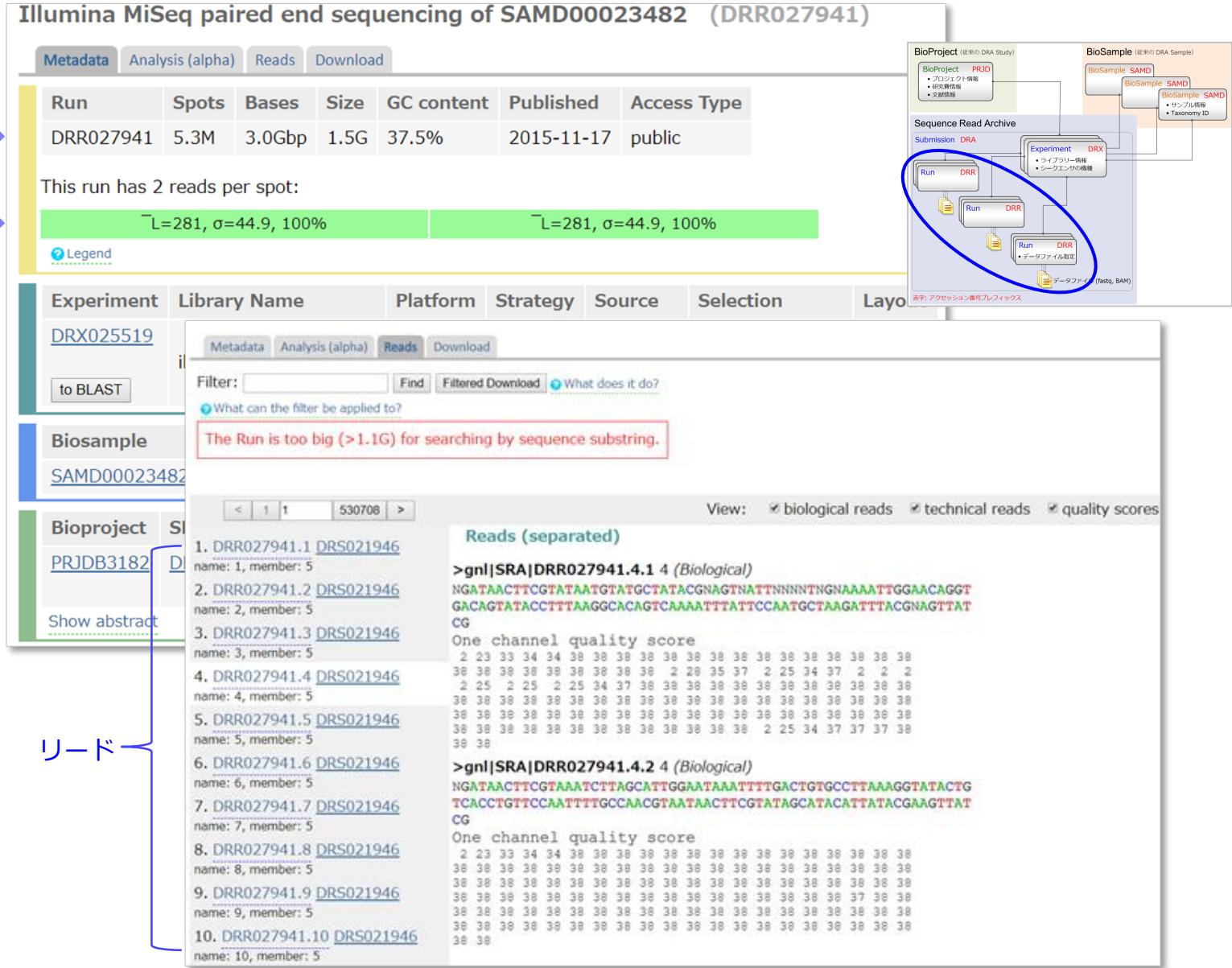
Run	# of Spots	# of Bases	Size	Published
DRR027941	5,307,079	3G	1.4Gb	2015-11-17
DRR027942	16,304,856	7.7G	4.7Gb	2015-11-17
DRR027940	5,904,000	3.4G	1.9Gb	2015-11-17

Run []



NCBI Experiment: <https://www.ncbi.nlm.nih.gov/sra/DRX025519>

SRA Run



NCBI Run: <https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=DRR027941>

Run 情報 →

リード構成 →

リード →

✓ リードデータのコンテナー

✓ リードには連番をアサイン

✓ Run ファイルは更新不可

✓ アクセッション番号プレフィックス

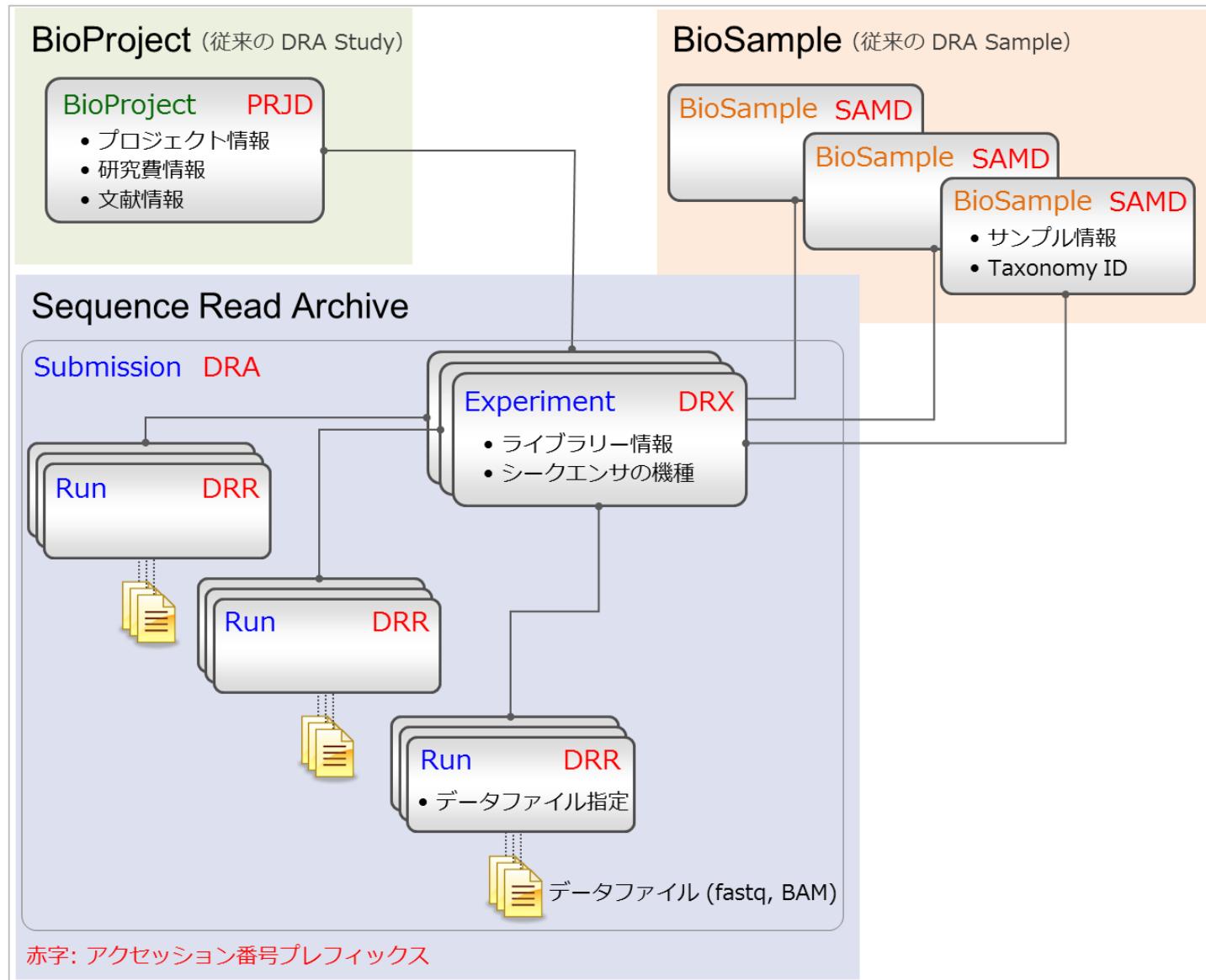
DRR (DDBJ)/ERR (EBI)/SRR (NCBI)

NCBI Run: <https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=DRR027941>

2018年8月29日

AJACS 番町 1

2. Sequence Read Archive (SRA)

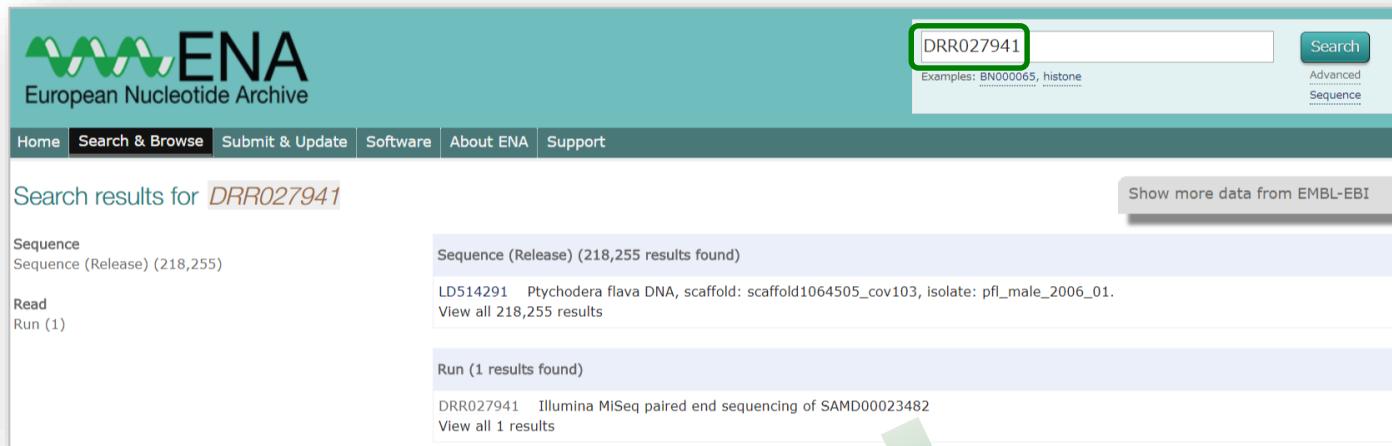


3. EBI/NCBI/DDBJ 各データベースの 使い方と特徴

EBI European Nucleotide Archive (ENA)

この章の内容を基にした統合 TV:

ENA (European Nucleotide Archive) を使って NGSデータとそのメタデータを効率よく検索する
<http://togotv.dbcls.jp/20180801.html>



Search results for *DRR027941*

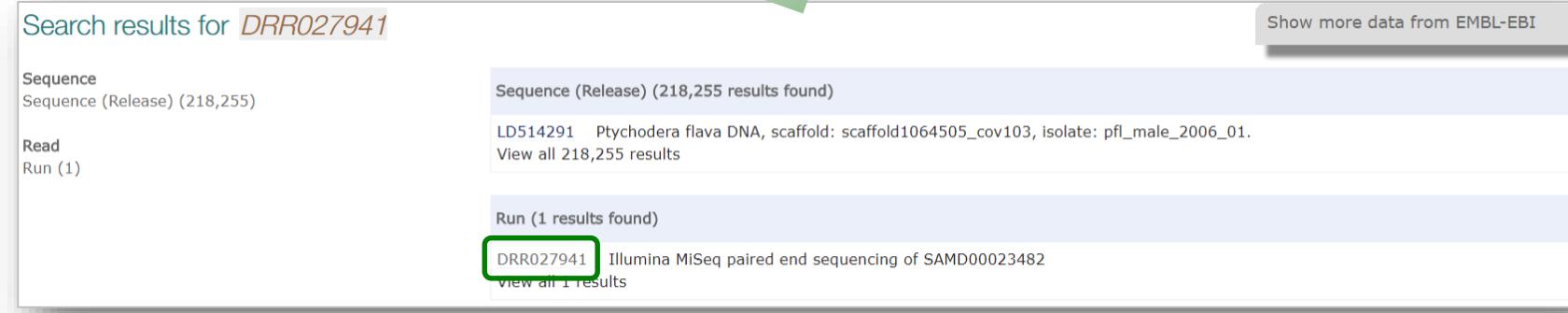
Sequence
Sequence (Release) (218,255)

Read
Run (1)

Sequence (Release) (218,255 results found)
LD514291 Pythodera flava DNA, scaffold: scaffold1064505_cov103, isolate: pfl_male_2006_01.
View all 218,255 results

Run (1 results found)
DRR027941 Illumina MiSeq paired end sequencing of SAMD00023482
View all 1 results

<https://www.ebi.ac.uk/ena/data/search?query=DRR027941>



Search results for *DRR027941*

Sequence
Sequence (Release) (218,255)

Read
Run (1)

Sequence (Release) (218,255 results found)
LD514291 Pythodera flava DNA, scaffold: scaffold1064505_cov103, isolate: pfl_male_2006_01.
View all 218,255 results

Run (1 results found)
DRR027941 Illumina MiSeq paired end sequencing of SAMD00023482
View all 1 results

✓ Run アクセッション番号で検索

ENA Search: 表形式でのプレゼンテーション

対象オブジェクトの情報

Run: DRR027941
Illumina MiSeq paired end sequencing of SAMD00023482

View: XML Download: XML

Submitting Centre OIST	Platform ILLUMINA	Model Illumina MiSeq	Read Count 5,307,079	Base Count 2,982,184,002
Library Layout PAIRED	Library Strategy WGS	Library Source GENOMIC	Library Selection size fractionation	Library Name illumina_mp_miseq_5k
Broker Name DDBJ				
Navigation Read Files				
This table contains the files for run DRR027941 Bulk Download Files ⚠ If the downloader app doesn't open, please try opening it directly. Download: 1 - 1 of 1 results in TEXT				
Hide Select columns <input checked="" type="checkbox"/> Study accession <input type="checkbox"/> Secondary study accession <input checked="" type="checkbox"/> Experiment accession <input type="checkbox"/> Run accession <input checked="" type="checkbox"/> Scientific name <input type="checkbox"/> Instrument platform <input type="checkbox"/> Paired nominal length <input type="checkbox"/> Library strategy <input type="checkbox"/> Library selection <input type="checkbox"/> Base count <input type="checkbox"/> First public <input type="checkbox"/> Experiment title <input type="checkbox"/> Submitter's study name <input type="checkbox"/> Submitter's run name <input type="checkbox"/> Fastq md5 <input type="checkbox"/> Submitter's sample name <input type="checkbox"/> Submitted bytes <input type="checkbox"/> Sample title <input type="checkbox"/> Submitted files (Galaxy) <input type="checkbox"/> First public <input checked="" type="checkbox"/> Submitted files (FTP) <input type="checkbox"/> Secondary sample accession <input checked="" type="checkbox"/> NCBI SRA file (FTP) <input type="checkbox"/> Experiment accession <input type="checkbox"/> CRAM Index files (Aspera) <input type="checkbox"/> Run accession <input type="checkbox"/> Sample title <input type="checkbox"/> Instrument model <input type="checkbox"/> Sample accession <input type="checkbox"/> Library layout <input type="checkbox"/> Submission accession <input type="checkbox"/> Read count <input type="checkbox"/> Instrument model <input type="checkbox"/> Last updated <input type="checkbox"/> Library strategy <input type="checkbox"/> Submitters's experiment name <input type="checkbox"/> Base count <input type="checkbox"/> FASTQ files (FTP) <input type="checkbox"/> Experiment title <input type="checkbox"/> Submitted MD5 <input type="checkbox"/> Submitter's run name <input type="checkbox"/> Submitted format <input type="checkbox"/> Submitter's sample name <input type="checkbox"/> NCBI SRA file (Aspera) <input type="checkbox"/> First public <input type="checkbox"/> CRAM Index files (Galaxy) <input type="checkbox"/> Secondary sample accession <input type="checkbox"/> Paired nominal sdev				
Showing results 1 - 1 of 1 results				
Study accession PRJDB3182	Sample accession SAMD00023482	Secondary sample accession DRS021946	Experiment accession DRX025519	Run accession DRR027941
Tax ID 63121	Scientific name Ptychoderidae	Instrument model Illumina MiSeq	Library layout PAIRED	FASTQ files (FTP) File 1 File 2
				FASTQ files (Galaxy) File 1 File 2
				Submitted files (FTP) File 1
				Submitted files (Galaxy) File 1
				NCBI SRA file (FTP) File 1
				NCBI SRA file (Galaxy) File 1
				CRAM Index files (FTP)
				CRAM Index files (Galaxy)

<https://www.ebi.ac.uk/ena/data/view/DRR027941>

- ✓ 表形式で関連オブジェクトとの対応関係が分かりやすい
- ✓ 表に含める項目を選択することができる

ENA Search: Studyで全体を一望

Study: PRJDB3182
Whole Genome Shotgun Sequencing of an Hawaiian Acornworm, Ptychodera flava.

View: Project XML Study XML

Name: Ptychodera flava Submitting Centre

Secondary accession(s): DRP002816

Description: Whole Genome Shotgun Sequencing of an Hawaiian Acornworm, Ptychodera flava. All original genome sequences were collected from Oahu Island in Hawaii at 2006 Dec. Ptychodera flava is an acornworm which shows indirect development. The accession number for sequences.

Lineage: Eukaryota, Metazoa, Hemichordata, Entoproctida, Ptychodera

Navigation: Read Files

Bulk Download Files (If the downloader app doesn't open, please try using Firefox to launch)

Download: 1 - 27 of 27 results in TEXT

Select columns

Showing results 1 - 10 of 27 results

Study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	Library layout	FASTQ files (FTP)	FASTQ files (Galaxy)	Submitted files (FTP)	Submitted files (Galaxy)	NCBI SRA files (FTP)	bioRxiv
PRJDB3182	SAMD00023482	DRS021946	DRX025509	DRR027930	63121	Ptychodera flava	454 GS FLX	SINGLE	File 1	File 1			File 1	
PRJDB3182	SAMD00023482	DRS021946	DRX025510	DRR027931	63121	Ptychodera flava	454 GS FLX Titanium	SINGLE	File 1	File 1			File 1	
PRJDB3182	SAMD00023482	DRS021946	DRX025511	DRR027932	63121	Ptychodera flava	454 GS FLX+	SINGLE	File 1	File 1			File 1	
PRJDB3182	SAMD00023482	DRS021946	DRX025512	DRR027933	63121	Ptychodera flava	Illumina MiSeq	PAIRED	File 1	File 1	File 1	File 1	File 1	
PRJDB3182	SAMD00023482	DRS021946	DRX025513	DRR027934	63121	Ptychodera flava	Illumina HiSeq 2000	PAIRED	File 1	File 1	File 1	File 1	File 1	
PRJDB3182	SAMD00023482	DRS021946	DRX025514	DRR027935	63121	Ptychodera flava	Illumina MiSeq	PAIRED	File 1	File 1	File 1	File 1	File 1	
PRJDB3182	SAMD00023482	DRS021946	DRX025515	DRR027936	63121	Ptychodera flava	Illumina MiSeq	PAIRED	File 1	File 1	File 1	File 1	File 1	
PRJDB3182	SAMD00023482	DRS021946	DRX025516	DRR027937	63121	Ptychodera flava	Illumina MiSeq	PAIRED	File 1	File 1	File 1	File 1	File 1	
PRJDB3182	SAMD00023482	DRS021946	DRX025517	DRR027938	63121	Ptychodera flava	Illumina MiSeq	PAIRED	File 1	File 1	File 1	File 1	File 1	
PRJDB3182	SAMD00023482	DRS021946	DRX025518	DRR027939	63121	Ptychodera flava	Illumina MiSeq	PAIRED	File 1	File 1	File 1	File 1	File 1	

Prev Next Go to page: 1 Go

<https://www.ebi.ac.uk/ena/data/view/PRJDB3182>

Study 配下の全オブジェクト

Java ツールでデータファイルをまとめてダウンロードできる

ENAFTP Downloader

Local Download Folder: C:\ena

Create Subfolder for each accession

Remote Files

FASTQ Submitted SRA

Download Accession Name

DRR027930 DRR027930.fastq.gz

DRR027931 DRR027931.fastq.gz

DRR027932 DRR027932.fastq.gz

DRR027933 DRR027933_1.fastq.gz

DRR027933_2.fastq.gz

DRR027934 DRR027934_1.fastq.gz

DRR027934_2.fastq.gz

DRR027935 DRR027935_1.fastq.gz

DRR027935_2.fastq.gz

DRR027936 DRR027936_1.fastq.gz

DRR027936_2.fastq.gz

DRR027937 DRR027937_1.fastq.gz

4 FASTQ files selected. Total size: 11

Select All Back

BioProject (PRJDB)

BioSample (SAMD)

BioSample (SAMD)

Sequence Read Archive

Submission (DRA)

Experiment (DRX)

Run (DRR)

Run (DRR)

Run (DRR)

データファイル配下

データファイル (Fastq, BAM)

表示: アクセシジョン番号プレフィックス

Red arrows point from the highlighted 'BioProject' and 'BioSample' sections in the ENAFTP Downloader interface to their corresponding components in the BioProject diagram.

✓ Study (BioProject) で全体を一望できる

ENA Advanced Search

The screenshot shows the European Nucleotide Archive (ENA) search interface. At the top, there's a navigation bar with links for Services, Research, Training, and About us. Below this is the ENA logo and a search bar containing the placeholder text "Example: BN000065, histone". To the right of the search bar is a "Search" button and an "Advanced Sequence" link, which is circled in green. A large green arrow points down to the "Advanced" search section. This section includes a radio button for "Advanced" (which is selected), another for "Upload accession", and a "Search query" input field. Below the input field are "Search" and "Clear" buttons. To the right of the search area is a "Help" link. On the left side of the page, there are two search sections: "Text Search" and "Sequence Search". The "Text Search" section has a search input field with examples like "BN000065, histone", a "Search" button, and a "Advanced search" link. The "Sequence Search" section has a search input field with examples like "Enter or paste a nucleotide sequence or accession number", a "Search" button, and a "Advanced search" link. At the bottom left, the URL "https://www.ebi.ac.uk/ena/data/warehouse/search" is displayed. On the right side, there's a "Select domain:" section with a list of options: Assembly, Sequence, Contig set, Coding, Non-coding, Read, Analysis, Study, Sample, Environmental, Taxon, and Marker. Most of these options have radio buttons next to them.

ENBL-EBI

ENA European Nucleotide Archive

Home Search & Browse Submit & Update Software About ENA Support

European Nucleotide Archive

The European Nucleotide Archive (ENA) provides a comprehensive record of the world's nucleotide sequence information, covering raw sequencing data, sequence assembly information and functional annotation. More ENA

Access to ENA data is provided through the browser, through search tools, large scale file download and through API.

Text Search

Examples: BN000065, histone

Search

Advanced search

Sequence Search

Enter or paste a nucleotide sequence or accession number

Search

Advanced search

Popular

Advanced Upload accession

Search query

Help

Select domain:

- Assembly
- Sequence
- Contig set
- Coding
- Non-coding
- Read
- Analysis
- Study
- Sample
- Environmental
- Taxon
- Marker

<https://www.ebi.ac.uk/ena/data/warehouse/search>

ENA Advanced Search: ENA 用語

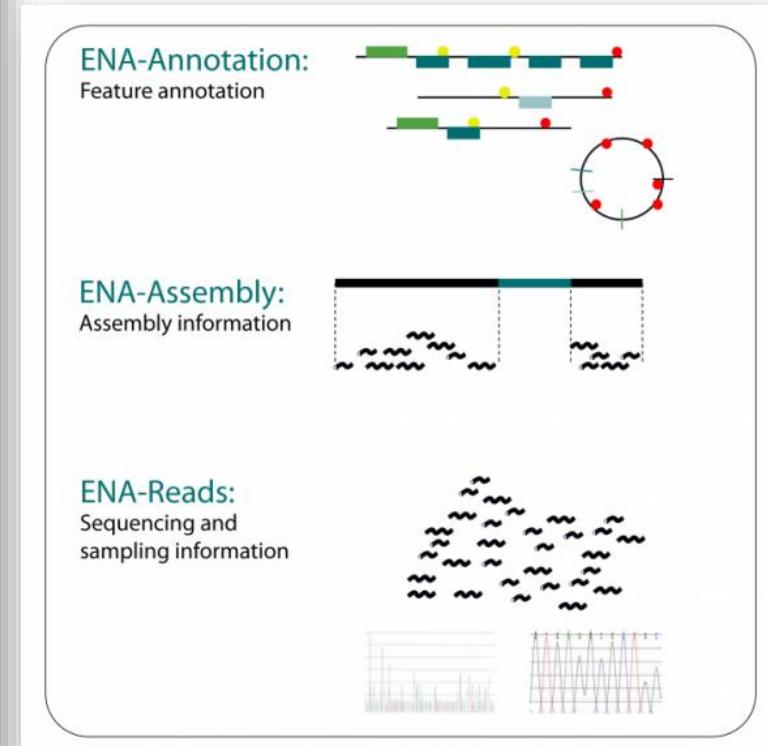
Advanced Upload accession

Search query Help

Search Clear

Select domain:

- Assembly ← Assembled sequences (Genomes)
- Sequence ← Annotated sequences
- Contig set
- Coding
- Non-coding
- Read ← SRA
- Analysis
- Study ← BioProject
- Sample ← BioSample
- Environmental
- Taxon ← Taxonomy: 生物分類
- Marker



<https://www.ebi.ac.uk/training/online/course/european-nucleotide-archive-quick-tour/what-european-nucleotide-archive>

- ✓ ENA domain とデータ種別・データベースとの対応関係を理解する

ENA Advanced Search: Read domain

Select domain:

- Assembly
- Sequence
- Contig set
- Coding
- Non-coding
- Read ← SRA
- Analysis
- Study
- Sample
- Environmental
- Taxon
- Marker

Select search conditions:

Taxonomy and related

Taxon name =
 Include subordinate taxa

Methodology

Library layout =
Instrument model =
Instrument platform =
Library selection =
Library strategy =
Library source =
Paired nominal length =
Paired nominal sdev =

Sequenced molecule

Base count =
Read count =

Text descriptions

Library name =
Experiment title =
Study title =

Database record

Study accession =
Secondary study accession =
Parent study accession =
Run accession =
Experiment accession =
Sample accession =
Secondary sample accession =
Submission accession =
Center name =
First public =

First created =

Submitter's study name =
Submitter's experiment name =
Submitter's sample name =
Submitter's run name =
Submitted format =
Broker name =

Calendars (Top right):
2018 Jan (with date 8 highlighted)
2018 Feb (with date 1 highlighted)

Calendars (Bottom right):
2018 Jan (with date 8 highlighted)
2018 Feb (with date 1 highlighted)

✓ Experiment, Run, Submission, Taxonomy の様々な項目で検索ができる

ENA Advanced Search: 例 1

- Illumina リード [instrument_platform="ILLUMINA"]
- ペアードシークエンス [library_layout="PAIRED"]
- 初回公開日が2017年1月1日以降 [first_public>=2017-01-01]

https://www.ebi.ac.uk/ena/data/warehouse/search?query=%22library_layout=%22PAIRED%22%20AND%20instrument_platform=%22ILLUMINA%22%20AND%20library_source=%22GENOMIC%22%20AND%20first_public%3E=2017-01-01%22&domain=read

Search query: library_layout="PAIRED" AND instrument_platform="ILLUMINA" AND library_source="GENOMIC" AND first_public>=2017-01-01

Search results for library_layout="PAIRED" AND instrument_platform="ILLUMINA" AND library_source="GENOMIC" AND first_public>=2017-01-01

Read Experiment (376,038) Run (388,671)

Study Study (8,709)

Records Reports Run (388,671 results found)

Download: 1 - 100000 of 388,671 results in XML

Showing results 1 - 10 of 388,671 results

DRR014457 Illumina MiSeq paired end sequencing; Whole genome sequencing of... more

DRR014565 Illumina Genome Analyzer IIx sequencing; sequencing of V. cholera... more

DRR014566 Illumina Genome Analyzer IIx sequencing: sequencing of V. cholera... more

DRR014567 Illumina Genome Analyzer IIx sequ... Navigation Read Files

DRR014568 Illumina Genome Analyzer IIx sequ...

DRR014569 Illumina Genome Analyzer IIx sequ...

DRR014570 Illumina Genome Analyzer IIx sequ...

DRR014882 Illumina HiSeq 2000 sequencing; L...

DRR016074 Illumina HiSeq 2000 sequencing; L...

DRR016075 Illumina HiSeq 2000 sequencing; B...

Download: 1 - 1 of 1 results Select columns

Showing results 1 - 1 of 1 results

Study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	Library layout	FASTQ files (FTP)	FASTQ files (Galaxy)	Submitted files (FTP)	Submitted files (Galaxy)	NCBI SRA file (FTP)	NCBI SRA file (Galaxy)	CRAM Index file (FTP)	CRAM Index files (Galaxy)
PRJDB2461	SAMD00012124	DRS013165	DRX013402	DRR014882	287	Pseudomonas aeruginosa	Illumina Genome Analyzer II	PAIRED	File 1 File 2 File 3	File 1 File 2 File 3			File 1	File 1		

Reports で検索結果をファイルでダウンロードできる

- DRR014882.fastq.gz # ペア無しリード
- DRR014882_1.fastq.gz # ペアリード1 forward
- DRR014882_2.fastq.gz # ペアリード2 reverse

ENA Advanced Search: 例 2

- Oryza sativa と下位の分類群 [tax_tree(4530)]
- 転写産物シークエンス [library_source="TRANSCRIPTOMIC"]

[https://www.ebi.ac.uk/ena/data/warehouse/search?query=%22tax_tree\(4530\)%20AND%20library_source=%22TRANSCRIPTOMIC%22%22&domain=read](https://www.ebi.ac.uk/ena/data/warehouse/search?query=%22tax_tree(4530)%20AND%20library_source=%22TRANSCRIPTOMIC%22%22&domain=read)

Search results for *library_source="TRANSCRIPTOMIC" AND tax_tree(4530)*

Read
Experiment (4,911)
Run (5,487)

Study
Study (510)

Records Reports

Download: 1 - 5487 of 5,487 results in TEXT

Select columns

Showing results 61 - 70 of 5,487 results
Please note that only the first 250 results are shown and that the results may be sorted differently in subsequent queries for best possible search performance. To view all of the results use the download link above.

Study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	Library layout
PRJNA306090	SAMN04338659	SRS1208316	SRX1484755	SRR3015588	39947	Oryza sativa Japonica Group	Illumina HiSeq 2500	PAIRED
PRJNA357146	SAMN06133444	SRS1859894	SRX2422826	SRR5110440	4530	Oryza sativa	Illumina HiSeq 2500	SINGLE
PRJNA122435	SAMN00008235	SRS010012	SRX016116	SRR034606	4530	Oryza sativa	Illumina Genome Analyzer	SINGLE
					17	Oryza sativa Indica Group	Illumina HiSeq 2500	PAIRED
					9	Oryza sativa	Illumina Genome Analyzer	SINGLE
					20	Oryza sativa	Illumina HiSeq 2500	PAIRED
					19767	Oryza sativa Japonica Group	Illumina HiSeq 2000	SINGLE
					39947	Oryza sativa	Illumina HiSeq 2500	PAIRED
						Oryza sativa	Illumina HiSeq 3000	PAIRED
						Oryza sativa Japonica Group	Illumina HiSeq 2000	SINGLE

• Oryza sativa
• Oryza sativa Japonica Group
• Oryza sativa Indica Group
がヒットする

Prev Next Go to page: 7 Go

ENA Advanced Search: 例 3

- Homo sapiens [tax_eq(9606)]
- Illumina [instrument_platform="ILLUMINA"]
- WGS (Whole genome shotgun) [library_strategy="WGS"]
- 2億リード以上 [read_count>=200000000]

[https://www.ebi.ac.uk/ena/data/warehouse/search?query=%22tax_eq\(9606\)%20AND%20instrument_platform=%22ILLUMINA%22%20AND%20library_strategy=%22WGS%22%20AND%20read_count%3E=200000000%22&domain=read](https://www.ebi.ac.uk/ena/data/warehouse/search?query=%22tax_eq(9606)%20AND%20instrument_platform=%22ILLUMINA%22%20AND%20library_strategy=%22WGS%22%20AND%20read_count%3E=200000000%22&domain=read)

This table contains the files for run SRR1273409

Bulk Download Files (If the downloader app doesn't open, please try using Firefox to launch it.)

Download: 1 - 1 of 1 results in TEXT

Select columns

Showing results 1 - 1 of 1 results

Study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	Library layout	FASTQ files (FTP)	FASTQ files (Galaxy)	Submitted files (FTP)	Submitted files (Galaxy)	NCBI SRA file (FTP)	NCBI SRA file (Galaxy)	CRAM Index files (FTP)	CRAM Index files (Galaxy)
PRJNA265937	SAMN03178165	SR5747758	SRX534465	SRR1273409	9606	Homo sapiens	Illumina HiSeq 2000	PAIRED			Go to dbGap	Go to dbGap				

- dbGaP 由来のアクセス制限データを除外
 - + submitted_format!="PROTECTED" [submitted_format!="PROTECTED"]

[https://www.ebi.ac.uk/ena/data/warehouse/search?query=%22tax_eq\(9606\)%20AND%20instrument_platform=%22ILLUMINA%22%20AND%20library_strategy=%22WGS%22%20AND%20read_count%3E=200000000%20AND%20submitted_format!=%22PROTECTED%22%22&domain=read](https://www.ebi.ac.uk/ena/data/warehouse/search?query=%22tax_eq(9606)%20AND%20instrument_platform=%22ILLUMINA%22%20AND%20library_strategy=%22WGS%22%20AND%20read_count%3E=200000000%20AND%20submitted_format!=%22PROTECTED%22%22&domain=read)

ENA Advanced Search: 例4 Sampleから

- domain: Sample
- taxonomy “[metagenomes \(408169\)](#)” と下位の分類群 [tax_tree(408169)]
- depth が200メートル以上 [depth>=200]

現時点では SRA の EBI Sample のファセット検索はできない

[https://www.ebi.ac.uk/ena/data/warehouse/search?query=%22tax_tree\(408169\)%20AND%20depth%3E=200%22&domain=sample](https://www.ebi.ac.uk/ena/data/warehouse/search?query=%22tax_tree(408169)%20AND%20depth%3E=200%22&domain=sample)

Sample: SAMEA3597864

Hydrothermal plume 1

View: [XML](#)

Name MG2-Plm-1

Submitting Centre Max Planck Institute for Marine Microbiology

Organism marine metagenon

Geographic location (country and/or sea) Atlantic Ocean

Investigation type metagenome

Project name Menez Gwen diffuse fluids

Sequencing method pyrosequencing

Collection date 2010-09-19/2010-10-04

Environmental package water

Geographic location (latitude) 37.8445833

Geographic location (longitude) -31.5188333

Environment (biome) marine pelagic biome

Environment (feature) marine hydrothermal plume

Environment (material) ocean water

Depth 797

Lineage unclassified sequences, metagenomes, ecological metagenomes

Navigation Read Files Attributes

大西洋深海797mの熱水噴出孔の海洋メタゲノムサンプル

Submitting Centre Max Planck Institute for Marine Microbiology

Organism marine metagenon

Geographic location (country and/or sea) Atlantic Ocean

Investigation type metagenome

Project name Menez Gwen diffuse fluids

Sequencing method pyrosequencing

Collection date 2010-09-19/2010-10-04

Environmental package water

Geographic location (latitude) 37.8445833

Geographic location (longitude) -31.5188333

Environment (biome) marine pelagic biome

Environment (feature) marine hydrothermal plume

Environment (material) ocean water

Depth 797

This table contains the files for sample ERS905013

Bulk Download Files (If the downloader app doesn't open, please try using Firefox to launch it.)

Download: - of 1 results in [TEXT](#)

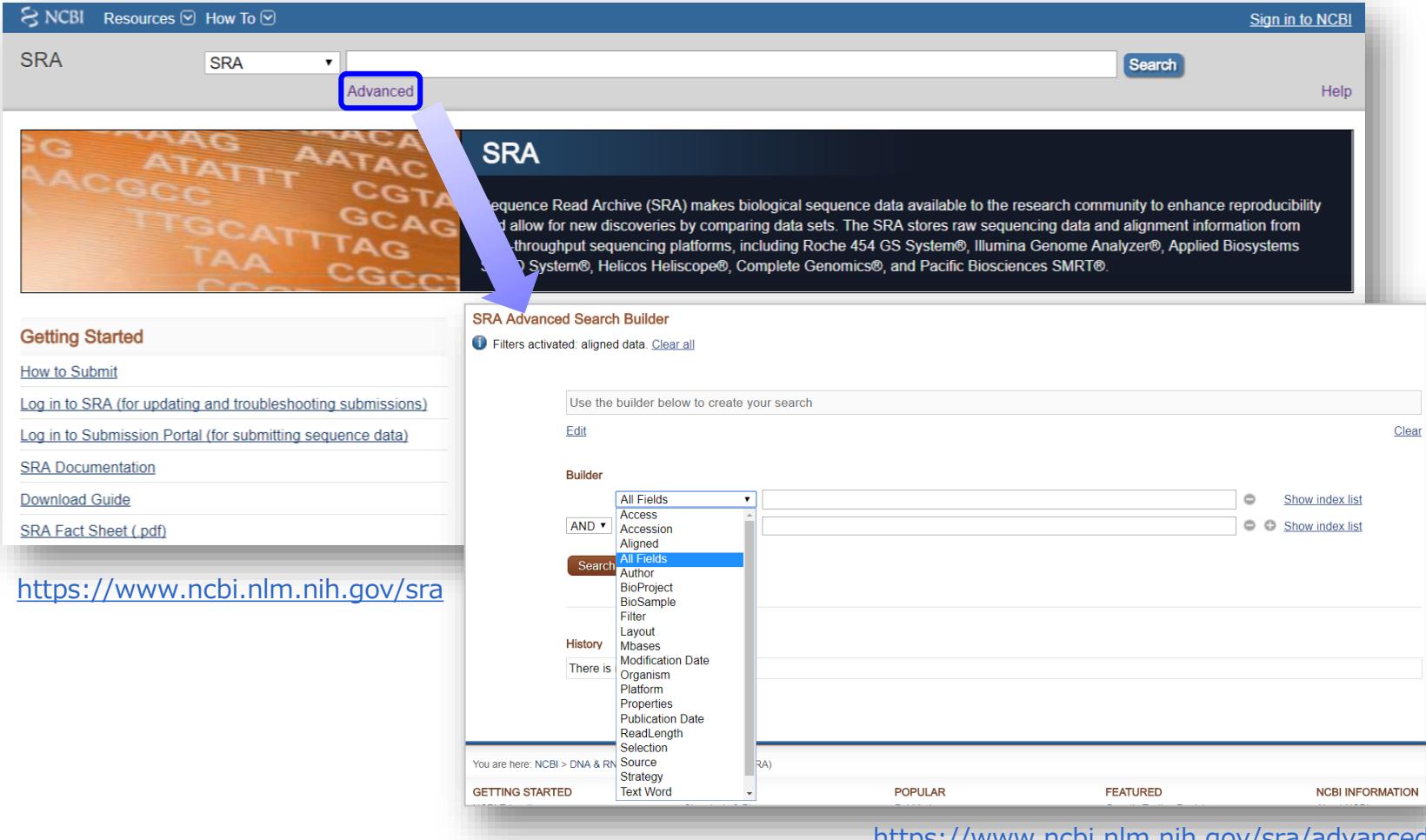
Select columns

Showing results 1 - 1 of 1 results

Study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	Library layout	FASTQ files (FTP)	FASTQ files (Galaxy)	Submitted files (FTP)	Submitted files (Galaxy)	NCBI SRA file (FTP)	NCBI SRA file (Galaxy)	CRAM Index files (FTP)	CRAM Index files (Galaxy)
PRJEB11362	SAMEA3597864	ERS905013	ERX1158322	ERR1078308	408172	marine metagenome	454 GS FLX+	SINGLE	File 1	File 1	SFF File 1	SFF File 1	File 1	File 1		

<https://www.ebi.ac.uk/ena/data/view/SAMEA3597864>

NCBI SRA



The screenshot shows the NCBI SRA search interface. At the top, there is a navigation bar with links for NCBI, Resources, How To, and Sign in to NCBI. Below the navigation bar, the search term "SRA" is entered in the search input field, and the "Advanced" button is highlighted with a blue box and a purple arrow pointing to it from the left.

The main content area displays the SRA homepage with a brief introduction about the Sequence Read Archive. Below the homepage, the "SRA Advanced Search Builder" is open. It includes a search bar with placeholder text "Use the builder below to create your search", an "Edit" link, and a "Clear" link. The search builder interface features a "Builder" section with dropdown menus for "All Fields" (selected), "AND" (selected), and "Search" (selected). A "History" section shows a search for "There is". The bottom of the builder interface includes a "GETTING STARTED" link, a "You are here: NCBI > DNA & RNA > SRA" breadcrumb trail, and category links for POPULAR, FEATURED, and NCBI INFORMATION. The URL <https://www.ncbi.nlm.nih.gov/sra> is shown at the bottom left, and the URL <https://www.ncbi.nlm.nih.gov/sra/advanced> is shown at the bottom right.

✓ Advanced で複数項目での絞り込み検索ができる

NCBI SRA: 高度なインテグレーション

アクセス制限
or
オープン
ファセット検索

The screenshot shows the NCBI SRA search results for "Homo sapiens". The search interface includes filters for access type (Controlled or Public), source (DNA/RNA), type (exome/genome), and other facets like aligned data. The main results list 12 entries, each detailing a metagenome run from LS454 with 2,787 to 4,285 spots and 1.2M to 1.9M bases. To the right, three sections are highlighted with blue arrows:

- Taxonomy ランキング**: Shows a tree view of top organisms, with Homo sapiens at the top (925082).
- BioProject ランキング**: Lists top bioprojects, including NIH Epigenomics Roadmap Initiative (1893) and Production ENCODE epigenomic (1493).
- 関連データベース レコード**: Displays records from BioSample, BioProject, dbGaP, and GEO Datasets.

<https://www.ncbi.nlm.nih.gov/sra/?term=Homo+sapiens>

NCBI SRA: Run selector で横に並べる

Ptychodera flava

Whole Genome Shotgun Sequencing of an Hawaiian Acornworm, Ptychodera flava.

Whole Genome Shotgun Sequencing of an Hawaiian Acornworm, Ptychodera flava. All original genomic library was produced from sperm spawned from one male individual which was sampled near Oahu island in Hawaii at 2006 Dec.
[More...](#)

Accession PRJDB3182

Data Type Genome sequencing and assembly

Scope Monoisolate

Organism Ptychodera flava [Taxonomy ID: 63121]
Eukaryota; Metazoa; Hemichordata; Enteropneusta; Ptychoderidae; Ptychodera; Ptychodera flava

Publications 1. Simakov O *et al.*, "Hemichordate genomes and deuterostome origins.", *Nature*, 2015 Nov 18;527(7579):459-65
2. Published online, DOI: 10.1038/nature16150

Submission Registration date: 24-Nov-2015
Okinawa Institute of Science and Technology

Related Resources • Marine Genomics Unit, OIST
• Marine Biological Laboratory (MBL), Graduate School of Science, Hiroshima University
• DRA002855
• BCFJ01000001-BCFJ01317432

Relevance Evolution

Project Data:

Resource Name	Number of Links
SEQUENCE DATA	
Nucleotide (total)	218256
WGS master	1
SRA Experiments	11
PUBLICATIONS	
PubMed	1
PMC	1
OTHER DATASETS	
BioSample	1
Assembly	1

Assembly details:

Assembly	Level	WGS	BioSample	Taxonomy
GCA_001465055.1	Scaffold	BCFJ00000000	SAMD00023482	Ptychodera flava

SRA Data Details

Parameter	Value
Data volume, Gbases	287
Data volume, Tbytes	0.23

<https://www.ncbi.nlm.nih.gov/bioproject/PRJDB3182>

Accession: PRJDB3182 ID: 302624

See Genome Information for Ptychodera flava

NAVIGATE ACROSS 3 additional projects are related by organism.

Summary ▾ 20 per page ▾ Send to: ▾

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

Links from BioProject

Items: 11

[Illumina MiSeq paired end sequencing of SAMD00023482](#)
1. 3 ILLUMINA (Illumina MiSeq) runs: 27.5M spots, 14.1G bases, 8Gb downloads
Accession: DRX025519

[Illumina MiSeq paired end sequencing of SAMD00023482](#)
2. 4 ILLUMINA (Illumina MiSeq) runs: 35.8M spots, 17.6G bases, 9.3Gb downloads
Accession: DRX025518

[Illumina MiSeq paired end sequencing of SAMD00023482](#)
3. 1 ILLUMINA (Illumina MiSeq) run: 6.9M spots, 3.9G bases, 2.1Gb downloads
Accession: DRX025517

[Illumina MiSeq paired end sequencing of SAMD00023482](#)
4. 1 ILLUMINA (Illumina MiSeq) run: 5.9M spots, 3.4G bases, 1.8Gb downloads
Accession: DRX025516

[Illumina MiSeq paired end sequencing of SAMD00023482](#)
5. 1 ILLUMINA (Illumina MiSeq) run: 10.7M spots, 6.1G bases, 3.4Gb downloads
Accession: DRX025515

[Illumina MiSeq paired end sequencing of SAMD00023482](#)
6. 1 ILLUMINA (Illumina MiSeq) run: 61.1M spots, 30.3G bases, 17.2Gb downloads
Accession: DRX025514

NCBI SRA: Run selector で横に並べる

Facets (left sidebar):

- Run
- Library name
- Platform
- MBases
- MBytes
- AvgSpotLen
- Experiment
- InsertSize
- Instrument
- LibraryLayout

Common Information (top right):

- Assay Type: WGS
- BioProject: PRJDB3182
- BioSample: SAMD00023482
- Center Name: OIST
- Consent: public
- LibrarySelection: size fractionation
- LibrarySource: GENOMIC
- LoadDate: 2015-12-29
- Organism: *Ptychoderida flava*
- ReleaseDate: 2015-11-17
- SRA Study: DRP002816
- Sample Name: DR5021946
- bioProject Id: PRJDB3182
- collection date: 2006-12-10
- description: Hawaiian Acornworm, *Ptychoderida flava*, Genomic DNA
- env biome: sea
- env feature: sand
- env material: sea water
- estimated size: 800,000,000
- geo loc name: USA: HI, Oahu
- isol growth condit: NA
- lat lon: NA
- num replicons: NA
- ploidy: haploid
- project name: *Ptychoderida flava* Genomic DNA
- propagation: NA
- ref biomaterial: NA
- sample name: *Ptychoderida flava* Genomic DNA
- sample title: *Ptychoderida flava* Genomic DNA

Run Information Table (bottom right):

Run	Library name	Platform	MBases	MBytes	AvgSpotLen	Experiment	InsertSize	Instrument	LibraryLayout
DRR027856	Illumina_pe_miseq_400	ILLUMINA	316	200	498	DRX025512	400	Illumina MiSeq	PAIRED
DRR027955	Illumina_pe_miseq_400	ILLUMINA	316	200	498	DRX025512	400	Illumina MiSeq	PAIRED
DRR027954	Illumina_pe_miseq_400	ILLUMINA	11,907	7,438	656	DRX025512	400	Illumina MiSeq	PAIRED
DRR027953	Illumina_pe_miseq_400	ILLUMINA	11,337	7,475	669	DRX025512	400	Illumina MiSeq	PAIRED
DRR027952	Illumina_pe_miseq_400	ILLUMINA	11,461	7,373	608	DRX025512	400	Illumina MiSeq	PAIRED
DRR027951	Illumina_pe_miseq_400	ILLUMINA	7,041	4,146	524	DRX025512	400	Illumina MiSeq	PAIRED
DRR027950	Illumina_pe_miseq_400	ILLUMINA	6,719	3,866	524	DRX025512	400	Illumina MiSeq	PAIRED
DRR027949	Illumina_pe_miseq_400	ILLUMINA	7,667	4,414	524	DRX025512	400	Illumina MiSeq	PAIRED
DRR027948	Illumina_pe_miseq_400	ILLUMINA	5,737	3,356	524	DRX025512	400	Illumina MiSeq	PAIRED
DRR027947	Illumina_pe_miseq_400	ILLUMINA	554	358	517	DRX025512	400	Illumina MiSeq	PAIRED
DRR027946	Illumina_pe_miseq_400	ILLUMINA	800	452	499	DRX025512	400	Illumina MiSeq	PAIRED
DRR027945	Illumina_mp_miseq_7k	ILLUMINA	8,441	5,073	479	DRX025512	7000	Illumina MiSeq	PAIRED
DRR027944	Illumina_mp_miseq_7k	ILLUMINA	4,730	2,462	469	DRX025518	7000	Illumina MiSeq	PAIRED
DRR027943	Illumina_mp_miseq_7k	ILLUMINA	1,707	864	550	DRX025518	7000	Illumina MiSeq	PAIRED
DRR027942	Illumina_mp_miseq_5k	ILLUMINA	7,359	4,796	473	DRX025519	5000	Illumina MiSeq	PAIRED
DRR027941	Illumina_mp_miseq_5k	ILLUMINA	2,844	1,436	561	DRX025519	5000	Illumina MiSeq	PAIRED
DRR027940	Illumina_mp_miseq_5k	ILLUMINA	3,210	1,911	570	DRX025519	5000	Illumina MiSeq	PAIRED
DRR027939	Illumina_mp_miseq_7k	ILLUMINA	1,896	1,132	563	DRX025518	7000	Illumina MiSeq	PAIRED
DRR027938	Illumina_mp_miseq_10k	ILLUMINA	3,713	2,102	563	DRX025517	10000	Illumina MiSeq	PAIRED
DRR027937	Illumina_mp_miseq_15k	ILLUMINA	3,243	1,823	574	DRX025516	15000	Illumina MiSeq	PAIRED
DRR027936	Illumina_mp_miseq_20k	ILLUMINA	5,815	3,472	570	DRX025515	20000	Illumina MiSeq	PAIRED
DRR027935	Illumina_mp_miseq_3k	ILLUMINA	28,880	17,609	495	DRX025514	3000	Illumina MiSeq	PAIRED
DRR027934	Illumina_Hiseq_400	ILLUMINA	110,885	75,168	202	DRX025513	400	Illumina HiSeq 2000	PAIRED
DRR027933	Illumina_pe_miseq_400	ILLUMINA	608	376	499	DRX025512	400	Illumina MiSeq	PAIRED
DRR027932	454_wgs_FLXplus	LS454	17,083	39,220	1151	DRX025511	0	454 GS FLX+	SINGLE
DRR027931	454_wgs_titanium	LS454	8,721	20,038	554	DRX025510	0	454 GS FLX Titanium	SINGLE
DRR027930	454_wgs_FLX	LS454	375	778	267	DRX025509	0	454 GS FLX	SINGLE

Download Options (bottom center):

- RunInfo Table
- Accession List
- RunTable
- AccessionList

Annotations:

- Run と参照オブジェクトの表 (Run table)
- ファセット検索 (Facet search)
- 共通情報 (Common information)
- Run 情報の表をファイルでダウンロード (Download Run information table)
- アクセッション番号リストをファイルでダウンロード (Download Accession list)

https://www.ncbi.nlm.nih.gov/Traces/study/?WebEnv=NCID_1_25540919_130.14.22.33_5555_1515566939_2113655420_0MetA0_S_HStore&query_key=16

NCBI SRA: aligned data の viewer 表示

Homo sapiens strain:(JPT) Japanese in Tokyo HapMap population (human)
Accession: PRJNA59845 ID: 59845
Exome sequencing of (JPT) Japanese in Tokyo, Japan HapMap population
Exome sequencing of (JPT) Japanese in Tokyo, Japan HapMap population

Accession	PRJNA59845
Data Type	Exome
Scope	Monosolate
Organism	Homo sapiens [Taxonomy ID: 9606] Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominoidea; Homo; Homo sapiens
Submission	Registration date: 2-Nov-2010 1000 Genomes Project - 1000genomes
Relevance	Medical

Project Data:

Resource Name	Number of Links
SRA Experiments	147
OTHER DATASETS	105
BioSample	

SRA Data Details

Parameter	Value
Data volume, Gbases	1.666
Data volume, Tbytes	1.14

aligned data で絞り込み

Links from BioProject
Items: 3

Filters activated: aligned data. [Clear all](#)

1 Illumina random exon sequencing of genomic DNA paired-end library 'Pond-314377' containing sample 'NA18939'
1 ILLUMINA (Illumina HiSeq 2000) run: 3.2M spots, 485.3M bases, 222.5Mb downloads
Accession: SRX655465

2 Illumina random exon sequencing of genomic DNA paired-end library 'Pond-314377' containing sample 'NA18939'
2 ILLUMINA (Illumina HiSeq 2000) runs: 23.6M spots, 3.6G bases, 1.6Gb downloads
Accession: SRX655436

3 Illumina random exon sequencing of genomic DNA paired-end library 'Pond-314377' containing sample 'NA18939'
3 ILLUMINA (Illumina HiSeq 2000) runs: 24.8M spots, 3.8G bases, 1.7Gb downloads
Accession: SRX655425

aligned data で絞り込み

染色体を選択

この Run を Sequence Viewer で表示

<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA59845>

Metadata Alignment Analysis (alpha) Reads Download

Alignment Reads Bases Fraction
Primary 6.3M 478.7Mbp 98.64%

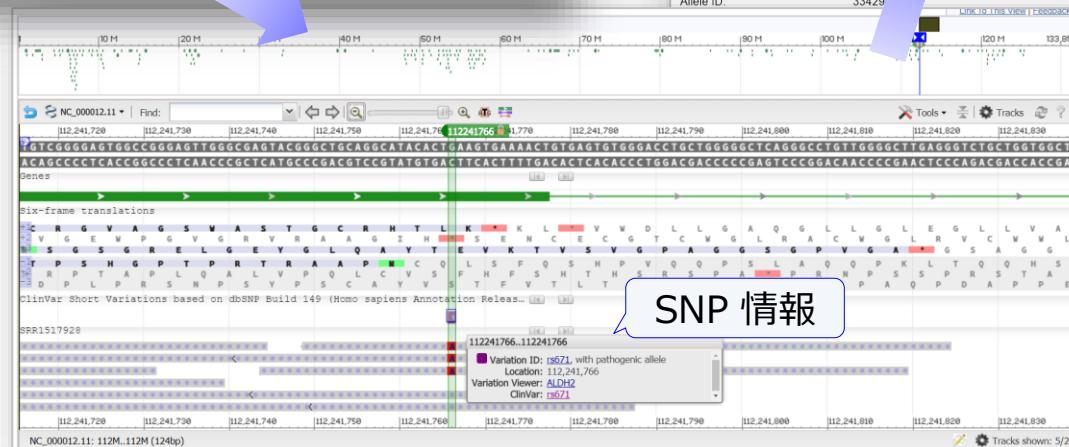
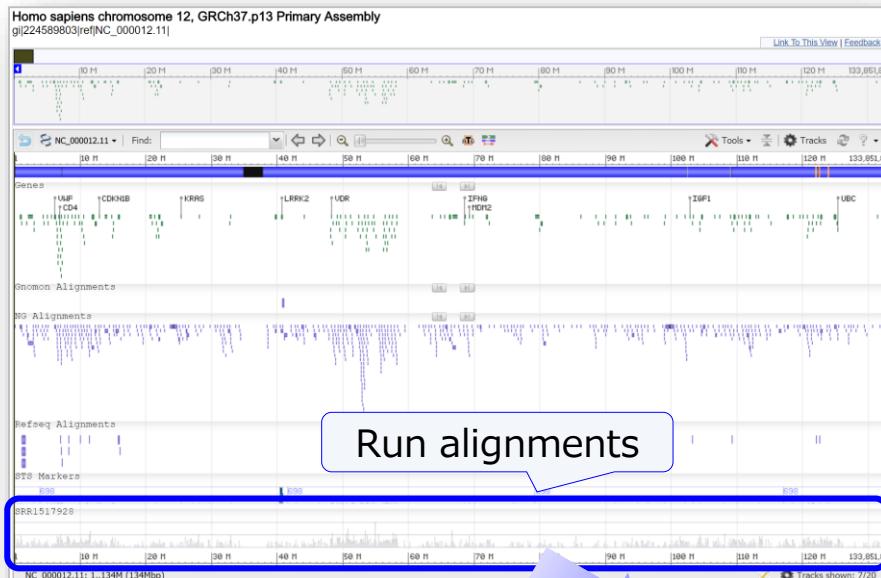
Reference 12
Homo sapiens chromosome 12, GRCh37.p13 Primary Assembly
[What does it do?](#)

View scope accession count in Sequence Viewer
this run SRR1517928 1
same experiment SRX655465 1
same sample SRS000715
same study SRP004076
all sra

Output this run in [FASTA](#) format to [Screen](#) [File](#)

<https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR1517928>

NCBI: Sequence viewer



Chr12: 112241766 (on Assembly GRCh37) <https://goo.gl/iX6xTJ>

<https://www.ncbi.nlm.nih.gov/clinvar/variation/18390/>

ClinVar

NM_000690.3(ALDH2):c.1510G>A (p.Glu504Lys)

Variation ID: 18390
Review status: (0/4) no assertion criteria provided

Interpretation

Clinical significance: Pathogenic, p.
Last evaluated: April 4, 2013
Number of submission(s): 5
Condition(s):

- Alcohol dependence
- Acute alcohol withdrawal
- Susceptibility
- Sublingual nitroglycerin tolerance
- Esophageal varices

See supporting ClinVar records

Allele(s)

NM_000690.3(ALDH2):c.1510G>A (p.Glu504Lys)

Allele ID: 33429

Go to: □

ClinVar: 変異の医学的な解釈
ALDH2, Glu504Lys (rs671)
GG: 高活性 日本人の56%
GA: 低活性 40%
AA: 非活性 4%

- ✓ アライメント, 遺伝子, 変異 etc を並べてブラウズし関連情報を辿ることができる

NCBI: 1000 Genomes Browser

1000 Genomes Browser Phase 3

Homo sapiens: GRCh37.p13 (GCF_000001405.25) Chr 12 (NC_000012.11): 112,241,734 - 112,241,799

ATTENTION: You are browsing the alignment and genotype data from the Phase 3 May 2013 call set.

Region ALDH2 NM_000690.3 Gene Transcript

Exons: click an exon above to zoom in, mouse over to see details

rs671

rs671 で検索

ALDH2, E504K (rs671)
 GG: 高活性 日本人の56%
 GA: 低活性 40%
 AA: 非活性 4%

Study: The 1000 Genomes Project (phase 3)

Show: Allele frequencies Genotype Display Options

Go To Selection	Scroll Region	112,241,495	112,241,552	112,241,553	112,241,597	112,241,613	112,241,614	112,241,624	112,241,648	112,241,720	112,241,730	112,241,747	112,241,766	112,241,771	112,241,795	112,241,804	112,241,811	112,241,833	112,241,903	112,241,943	112,241,974	112,242,002	112,242,023
		rs555019857	rs2106968	rs544019456	rs74938832	rs145526471	rs54526348	rs564363961	rs367711233	rs190914158	rs198724288	rs201939387	rs671	rs569388000	rs147178045	rs200975817	rs566179817	rs533526749	rs554905215	rs573421256	rs544247104	rs55902673	rs57757624
Populations / Samples		T=0.9998 A=0.0002	T=0.9436 A=0.0004	T=0.9995 C=0.0100	G=0.9990 C=0.0002	T=0.9998 T=0.0004	A=0.9998 G=0.0002	C=0.9996 T=0.0002	C=0.9996 A=0.0002	G=0.9998 T=0.0004	C=0.9998 T=0.0002	G=0.9998 T=0.0004	rs671	G=0.9643 A=0.0357	G=0.9994 A=0.0006	T=0.9974 C=0.026	G=0.9998 T=0.0002	G=0.9998 T=0.0002	C=0.9998 T=0.0002	G=0.9998 T=0.0004	G=0.9998 C=0.0002	G=0.9998 T=0.0002	G=0.9998 A=0.0004
FIN Finnish in Finland		T=1.0000 C=0.0000	T=0.9646 A=0.0354	T=1.0000 A=0.0000	G=1.0000 C=0.0000	T=1.0000 C=0.0000	A=0.9949 G=0.0051	C=1.0000 T=0.0000	C=1.0000 T=0.0000	G=1.0000 A=0.0000	C=1.0000 T=0.0000	G=1.0000 A=0.0000	G=1.0000 T=0.0000	G=1.0000 A=0.0000	T=1.0000 C=0.0000	T=1.0000 T=0.0000	G=1.0000 T=0.0000	C=1.0000 T=0.0000	G=1.0000 T=0.0000	G=1.0000 T=0.0000	G=1.0000 A=0.0101		
GBR British in England and S...		T=1.0000 C=0.0000	T=0.9121 A=0.0879	T=1.0000 A=0.0000	G=1.0000 C=0.0000	T=1.0000 C=0.0000	A=1.0000 G=0.0000	C=1.0000 T=0.0000	C=1.0000 A=0.0000	G=1.0000 T=0.0000	C=1.0000 A=0.0000	G=0.9880 T=0.0110	C=0.9945 T=0.0055	G=1.0000 A=0.0000	T=1.0000 C=0.0000	G=1.0000 T=0.0000	G=1.0000 T=0.0000	C=1.0000 T=0.0000	G=1.0000 T=0.0000	G=1.0000 A=0.0000			
GIH Gujarati Indian from Hou...		T=1.0000 C=0.0000	T=0.9126 A=0.0874	T=1.0000 A=0.0000	G=1.0000 C=0.0000	T=1.0000 C=0.0000	A=1.0000 G=0.0000	C=1.0000 T=0.0000	C=1.0000 A=0.0000	G=1.0000 T=0.0000	C=1.0000 A=0.0000	G=1.0000 T=0.0000	C=1.0000 T=0.0000	G=1.0000 A=0.0000	T=1.0000 C=0.0000	G=1.0000 T=0.0000	G=1.0000 T=0.0000	C=1.0000 T=0.0000	G=1.0000 T=0.0000	G=1.0000 A=0.0000			
GWD Gambian in Western Di...		T=1.0000 C=0.0000	T=0.6770 A=0.3230	T=1.0000 A=0.0000	G=0.9690 C=0.0310	T=1.0000 C=0.0000	A=1.0000 G=0.0000	C=1.0000 T=0.0000	C=1.0000 A=0.0000	G=1.0000 T=0.0000	C=1.0000 A=0.0000	G=1.0000 T=0.0000	C=1.0000 A=0.0044	G=0.9956 A=0.0044	T=1.0000 C=0.0000	G=1.0000 T=0.0000	G=1.0000 T=0.0000	C=1.0000 T=0.0000	G=1.0000 T=0.0000	G=1.0000 A=0.0000			
IBS Iberian Population in Spain		T=1.0000 C=0.0000	T=0.6832 A=0.1168	T=1.0000 A=0.0000	G=1.0000 C=0.0000	T=1.0000 C=0.0000	A=1.0000 G=0.0000	C=1.0000 T=0.0000	C=1.0000 A=0.0000	G=1.0000 T=0.0000	C=1.0000 A=0.0000	G=1.0000 T=0.0000	C=1.0000 T=0.0000	G=1.0000 A=0.0000	T=1.0000 C=0.0000	G=1.0000 T=0.0000	G=1.0000 T=0.0000	C=1.0000 T=0.0000	G=1.0000 T=0.0000	G=1.0000 A=0.0000			
ITU Indian Telugu from the UK		T=0.9951 C=0.0049	T=0.9265 A=0.0735	T=1.0000 A=0.0000	G=1.0000 C=0.0000	T=1.0000 C=0.0000	A=1.0000 G=0.0000	C=1.0000 T=0.0000	C=1.0000 A=0.0000	G=1.0000 T=0.0000	C=1.0000 A=0.0000	G=1.0000 T=0.0000	C=1.0000 T=0.0000	G=0.9853 A=0.0147	T=1.0000 C=0.0000	G=1.0000 T=0.0000	G=1.0000 T=0.0000	C=1.0000 T=0.0000	G=1.0000 T=0.0000	G=1.0000 A=0.0000			
JPT Japanese in Tokyo, Japan		T=1.0000 C=0.0000	T=0.4231 A=0.5769	T=1.0000 A=0.0048	G=1.0000 C=0.0000	T=1.0000 C=0.0000	A=1.0000 G=0.0000	C=1.0000 T=0.0000	C=1.0000 A=0.0000	G=1.0000 T=0.0000	C=1.0000 A=0.0000	G=0.7596 A=0.2404	C=1.0000 T=0.0000	G=1.0000 A=0.0000	T=1.0000 C=0.0000	G=1.0000 T=0.0000	G=1.0000 T=0.0000	C=1.0000 T=0.0000	G=1.0000 T=0.0000	G=1.0000 A=0.0000			
KHV Kinh in Ho Chi Minh City, Vietnam		T=1.0000 C=0.0000	T=0.4747 A=0.5253	T=1.0000 A=0.0000	G=1.0000 C=0.0000	T=1.0000 C=0.0000	A=1.0000 G=0.0000	C=1.0000 T=0.0000	C=1.0000 A=0.0000	G=1.0000 T=0.0000	C=1.0000 A=0.0000	G=0.8636 A=0.1364	C=1.0000 T=0.0000	G=1.0000 A=0.0000	T=1.0000 C=0.0000	G=1.0000 T=0.0000	G=1.0000 T=0.0000	C=1.0000 T=0.0000	G=1.0000 T=0.0000	G=1.0000 A=0.0000			

✓ 集団毎の頻度を調べることができる

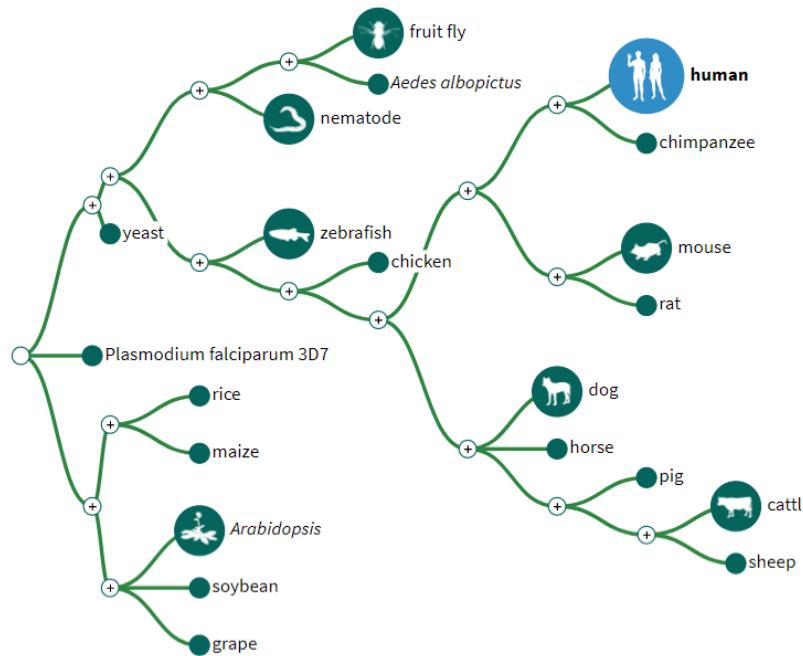
NCBI: Genome Data Viewer

Genome Data Viewer

GDV is a genome browser supporting the exploration and analysis of more than 540 eukaryotic RefSeq genome assemblies. [?](#)

Select organism

Homo sapiens (human)



Homo sapiens (human) genome

Search in genome:

Examples: TP53, Chr17:7667000-7669000, TSS34, DNA repair

Assembly: GRCh38.p11

[Browse genome](#) [BLAST genome](#)

Assembly details

Name	GRCh38.p11
RefSeq accession	GCF_000001405.37
GenBank accession	GCA_000001405.26
Download via FTP	RefSeq, GenBank
Submitter	Genome Reference Consortium
Level	Chromosome

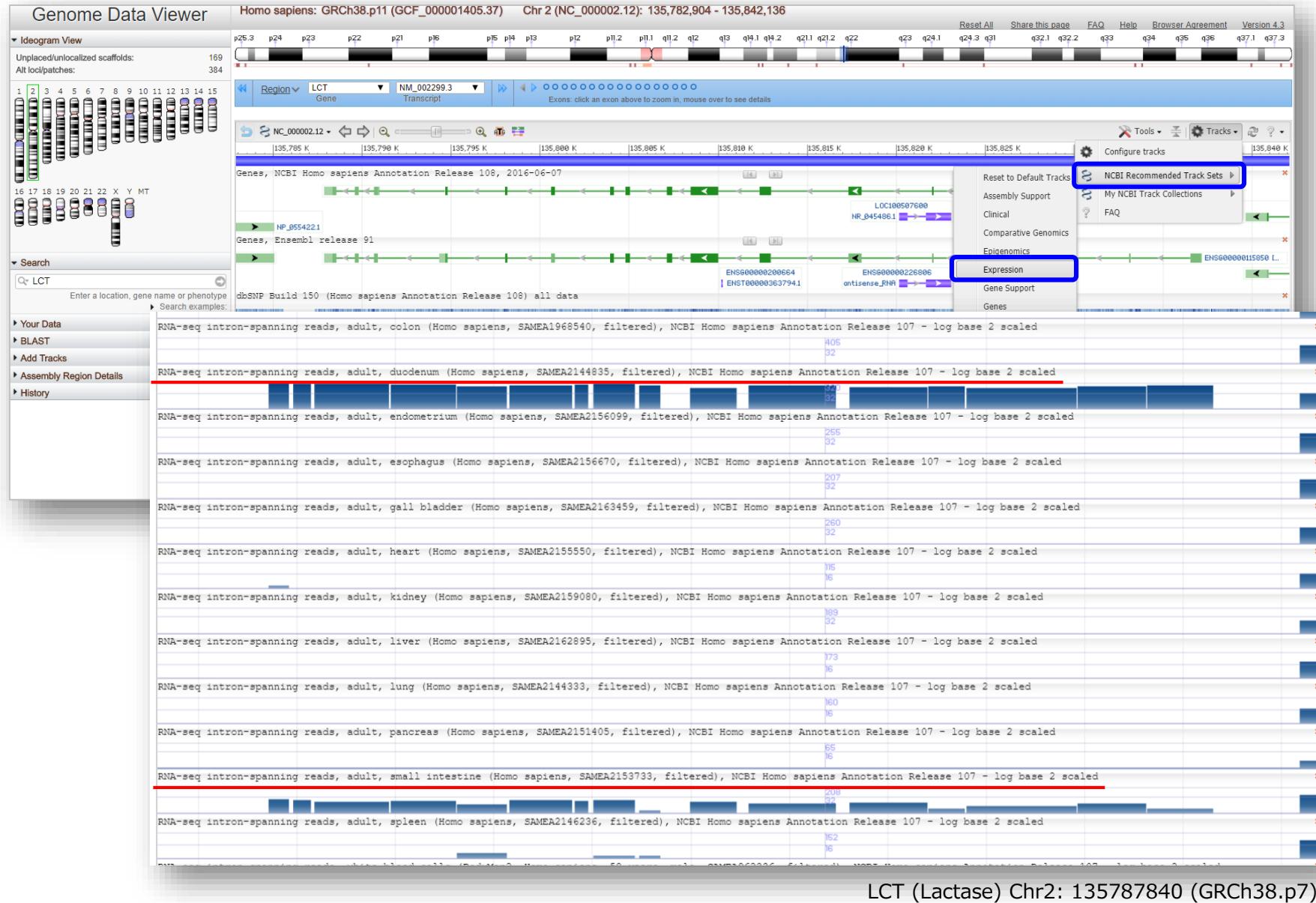
Annotation details

Annotation Release 108
Release date

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y

<https://www.ncbi.nlm.nih.gov/genome/gdv/>

NCBI: Genome Data Viewer (Expression)



NCBI: Gene (Expression)

Genome Data Viewer Homo sapiens: GRCh38.p11 (GCF_000001405.37) Chr 2 (NC_000002.12): 135,782,904 - 135,842,136

Reset All Share this page FAQ Help Browser Agreement Version 4.3

Ideogram View Unplaced/unlocalized scaffolds: 169 Alt loci/patches: 384

Region LCT NM_002299.3 Gene Transcript

Gene: LCT Title: lactase Location: complement(135,787,840..135,837,200) Length: 49,361 Merged features: 4

View GeneID: 3938 (LCT)
View HGNC: 9336
View HPRD: 04432
View MIM: 603202

LCT lactase [Homo sapiens (human)] Gene ID: 3938, updated on 14-Jan-2018

Summary

Official Symbol LCT provided by HGNC
Official Full Name lactase provided by HGNC
Primary source HGNC-HGNC:6530
See related Ensembl:ENSG00000115850 MIM:603202; Vega:OTTHUMG00000131738
Gene type protein coding
RefSeq status REVIEWED

Expression

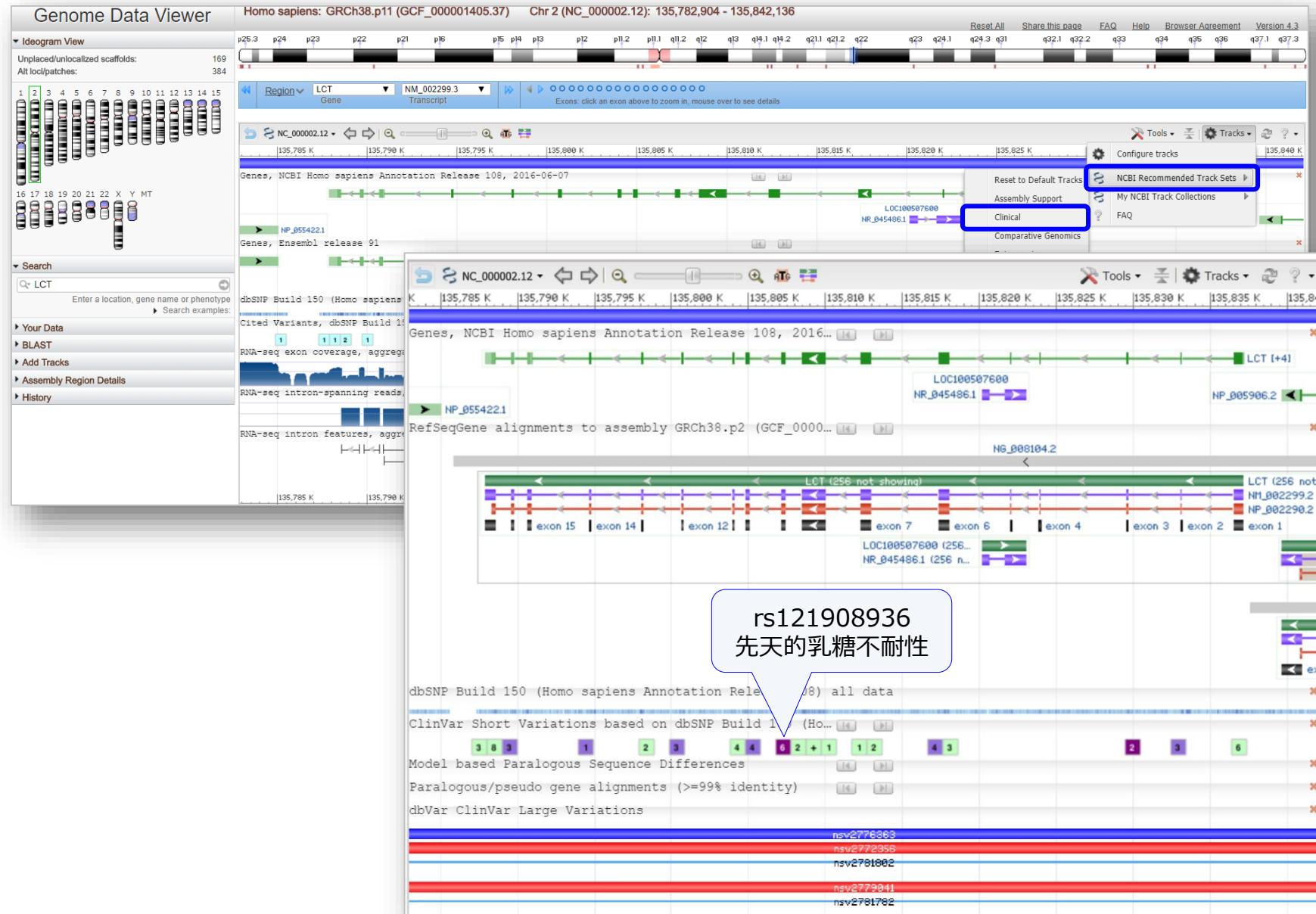
HPA RNA-seq normal tissues

- Project title: HPA RNA-seq normal tissues
- Description: RNA-seq was performed of tissue samples from 95 human individuals representing 27 different tissues in order to determine tissue-specificity of all protein-coding genes
- BioProject: PRJEB4337
- Publication: PMID 24309898
- Analysis date: Wed Jun 15 11:32:44 2016

Samples	RPKM
adrenal	0
appendix	0
bone marrow	0
brain	0
colon	0
duodenum	~120
endometrium	0
esophagus	0
fat	0
gall bladder	0
heart	0
kidney	0
liver	0
lung	0
lymph node	0
ovary	0
pancreas	0
placenta	0
prostate	0
salivary gland	0
skin	0
small intestine	~75
spleen	0
stomach	0
testis	0
thyroid	0
urinary bladder	0

<https://www.ncbi.nlm.nih.gov/gene/3938>

NCBI: Genome Data Viewer (Variation)



NCBI: SRA Run BLAST

<https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=DRR068403>

Illumina MiSeq sequencing of SAMD00058608 (DRR068403)

The screenshot illustrates the workflow for performing a BLAST search on a fecal microbiome sample from subject 516. The process starts with viewing the SRA run details and experiment metadata, then navigating to the BLAST interface.

Left Panel (SRA Run Details):

- Run: DRR068403, 7.3k Spots, 2.6Mbp, 1.6M Size, 52.9% GC content, Published 2016-11-17, Access Type public.
- This run has 2 reads per spot: L=353, 100%.
- Experiment: DRX062354 (to BLAST highlighted with a blue box).
- Biosample: SAMD00058608 (highlighted with a blue box).
- Bioproject: PRJDB4998 (highlighted with a green box), DRP003338: Fecal microbiota of Japanese healthy adults.
- Micromicrobiota of the fecal sample from subject 516 (highlighted with a blue box):
 - Identifiers: BioSample: SAMD00058608
 - Organism: human gut metagenome (unclassified sequences, metagenomes; organismal metagenomes)
 - Package: MIMS: metagenome/environmental, human-gut; version 4.0
 - Attributes (sample name: YB_516, collection date: 2013, environment biome: terrestrial biome, environment feature: human-associated habitat, environment material: feces, geographic location: Japan, host: Homo sapiens, host subject id: 516, latitude and longitude: not collected, project name: Fecal microbiota of Japanese healthy adult)
 - Description: Keywords: GSC:MixS;MIMS:4.0
 - BioProject: PRJDB4998 human gut metagenome (Retrieve all samples from this project)
 - Submission: Yakult Central Institute; 2016-11-17

Right Panel (BLAST Interface):

The BLAST interface shows the following steps:

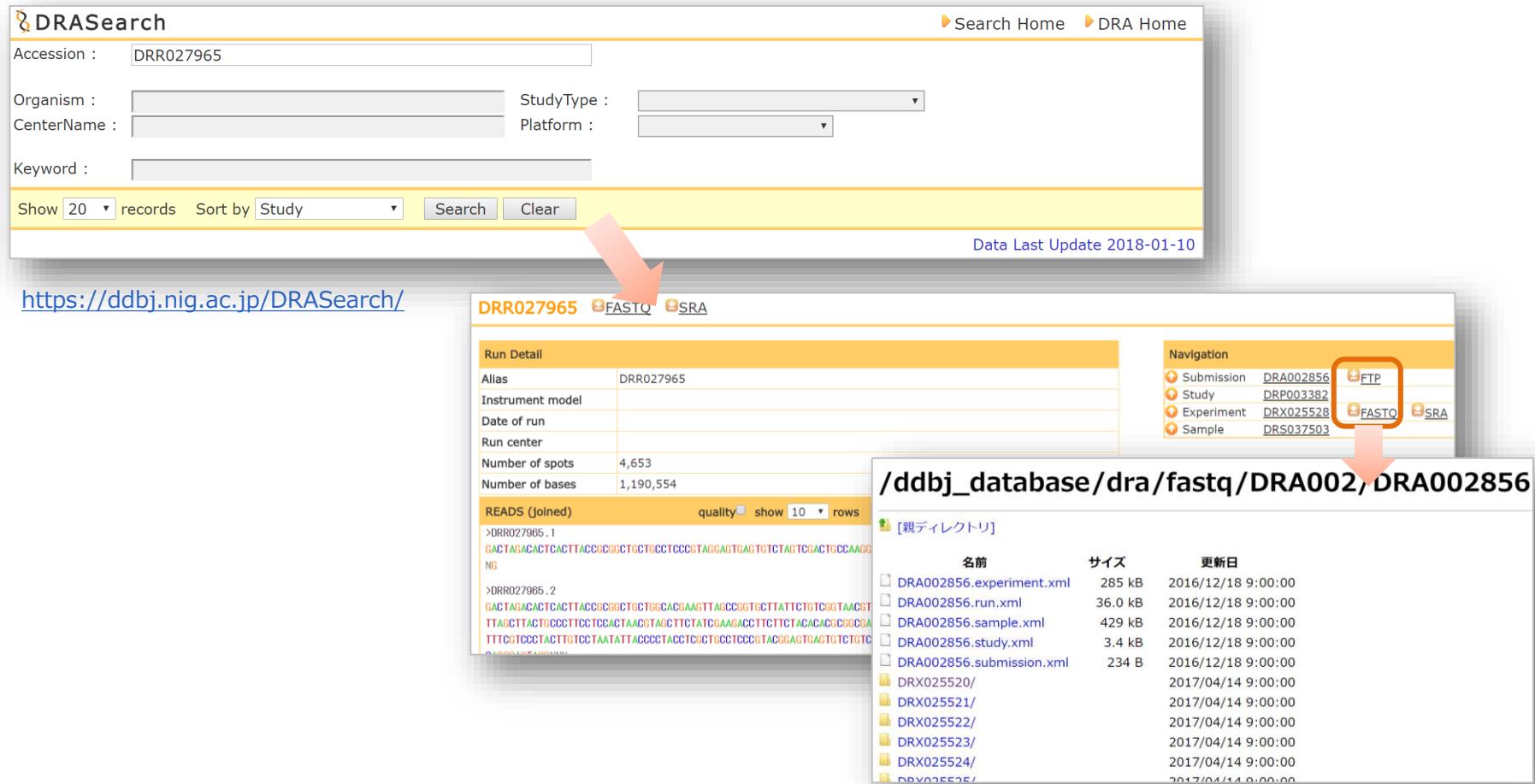
- Enter Query Sequence:** A purple arrow points from the 'to BLAST' button in the SRA panel to the 'Enter Query Sequence' field in the BLAST interface.
- Choose Search Set:** The search set is set to 'SRA Experiment set (SRX)'.
- Program Selection:** The program is set to 'BLAST'.
- Graphic Summary:** A histogram titled 'Distribution of the top 100 Blast Hits on 406 subject sequences' shows alignment scores. The x-axis ranges from 1 to 500, and the y-axis represents the number of subject sequences. A color key indicates scores: <40 (black), 40-50 (blue), 50-80 (green), 80-200 (magenta), and >=200 (red). Most hits are in the >=200 range.
- Descriptions:** A table lists sequences producing significant alignments, including their IDs (DR0002354), Max score, Total score, Query cover, E value, Ident, and Accession.

✓ ヒト腸内細菌メタゲノムリードに対してビフィズス菌属特異的 16S rRNA 配列で BLAST

NCBI SRA: Run Taxonomy Analysis

✓ リードの taxonomy クラスタリング結果（アルファ版）

DDBJ Sequence Read Archive



The screenshot shows the DRA Search interface. At the top, there are search fields for Accession (DRR027965), Organism, CenterName, StudyType, Platform, and Keyword. Below these are buttons for Show 20 records, Sort by Study, Search, and Clear. A red arrow points from the FASTQ and SRA download links in the main content area down to the corresponding links in the navigation sidebar. The sidebar also includes links for Submission (DRA002856), Study (DRP003382), Experiment (DRX025528), and Sample (DRS037503). A red box highlights the FASTQ and SRA links in both the main content and the sidebar.

✓ 国内であれば ftp で高速にダウンロードできる

aspera にも対応

<https://www.ddbj.nig.ac.jp/dra/faq.html#how-to-download-data>

DDBJ Search by DBCLS

DRA Home DDBJ flat file search

 DDBJ Search > sra

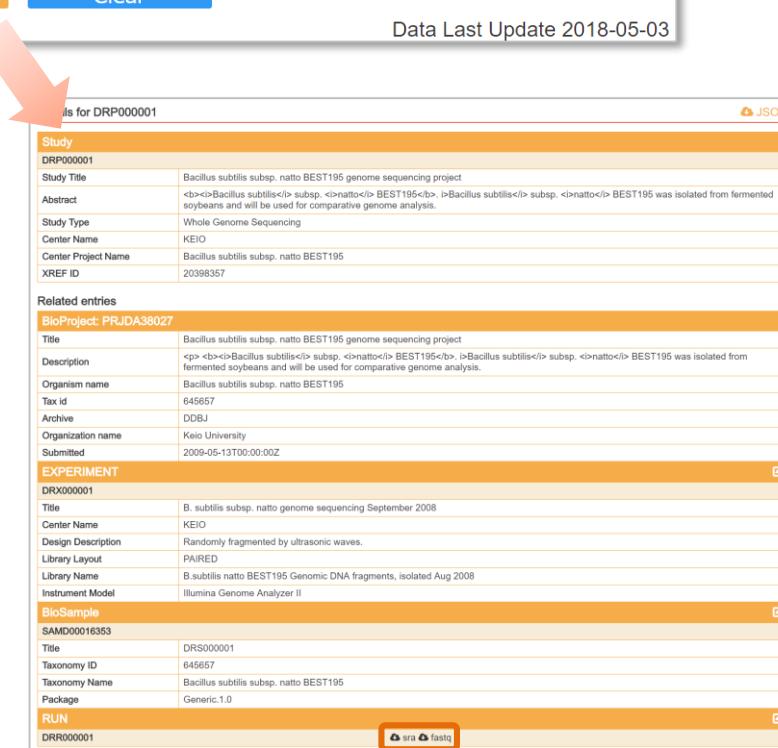
Keyword :

Accession :

Show 20 records Sort by ACCESSION

Data Last Update 2018-05-03

<http://sra.dbcls.jp/>



Results for DRP000001

Study

DRP000001	Bacillus subtilis subsp. natto BEST195 genome sequencing project
Abstract	<><>Bacillus subtilis</> subsp. <>natto</> BEST195</>. >Bacillus subtilis</> subsp. <>natto</> BEST195 was isolated from fermented soybeans and will be used for comparative genome analysis.
Study Type	Whole Genome Sequencing
Center Name	KEIO
Center Project Name	Bacillus subtilis subsp. natto BEST195
XREF ID	20398357

Related entries

BioProject: PRJDA38027	
Title	Bacillus subtilis subsp. natto BEST195 genome sequencing project
Description	<><>Bacillus subtilis</> subsp. <>natto</> BEST195</>. >Bacillus subtilis</> subsp. <>natto</> BEST195 was isolated from fermented soybeans and will be used for comparative genome analysis.
Organism name	Bacillus subtilis subsp. natto BEST195
Tax id	645657
Archive	DDBJ
Organization name	Keio University
Submitted	2009-05-13T00:00:00Z

EXPERIMENT

DRX000001	B. subtilis subsp. natto genome sequencing September 2008
Center Name	KEIO
Design Description	Randomly fragmented by ultrasonic waves.
Library Layout	PAIRED
Library Name	B. subtilis natto BEST195 Genomic DNA fragments, isolated Aug 2008
Instrument Model	Illumina Genome Analyzer II

BioSample

SAMD000016353	
Title	DRS000001
Taxonomy ID	645657
Taxonomy Name	Bacillus subtilis subsp. natto BEST195
Package	Generic.1.0

RUN

DRR000001	
-----------	--

✓ BioProject/BioSample/SRA を一覧できる

3. EBI/NCBI/DDBJ 各データベースの使い方と特徴



- ✓ メタデータの様々な項目での絞り込み検索ができる
- ✓ 表形式で全体像を把握しやすい
- ✓ ENA 以外の BioSample, ArrayExpress, EVA, EGA etc との連携は開発途上



- ✓ NCBI 各サービスとの高度なインテグレーション、特に ClinVar 等の医学系リソース
- ✓ アライメントデータは viewer でブラウズできる
- ✓ sra ファイルから fastq etc への変換が必要



- ✓ 国内であれば高速にデータファイルをダウンロードできる
- ✓ DDBJ Search by DBCLS は BioProject/BioSample/SRA を一覧できる