

次世代シークエンスデータベース の紹介と使い方

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



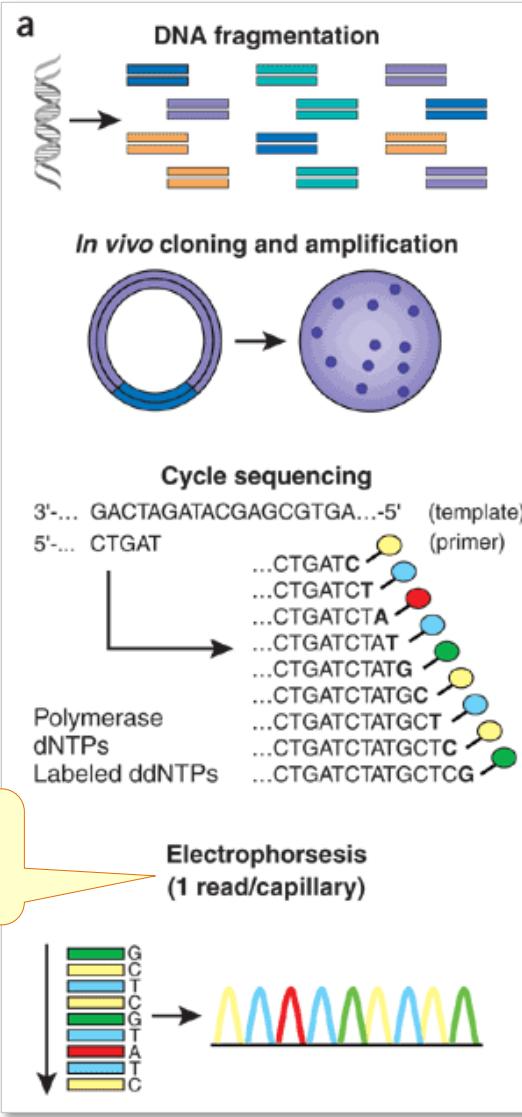
✓ 国立遺伝学研究所 DDBJ センター コーディネータ

講義の内容

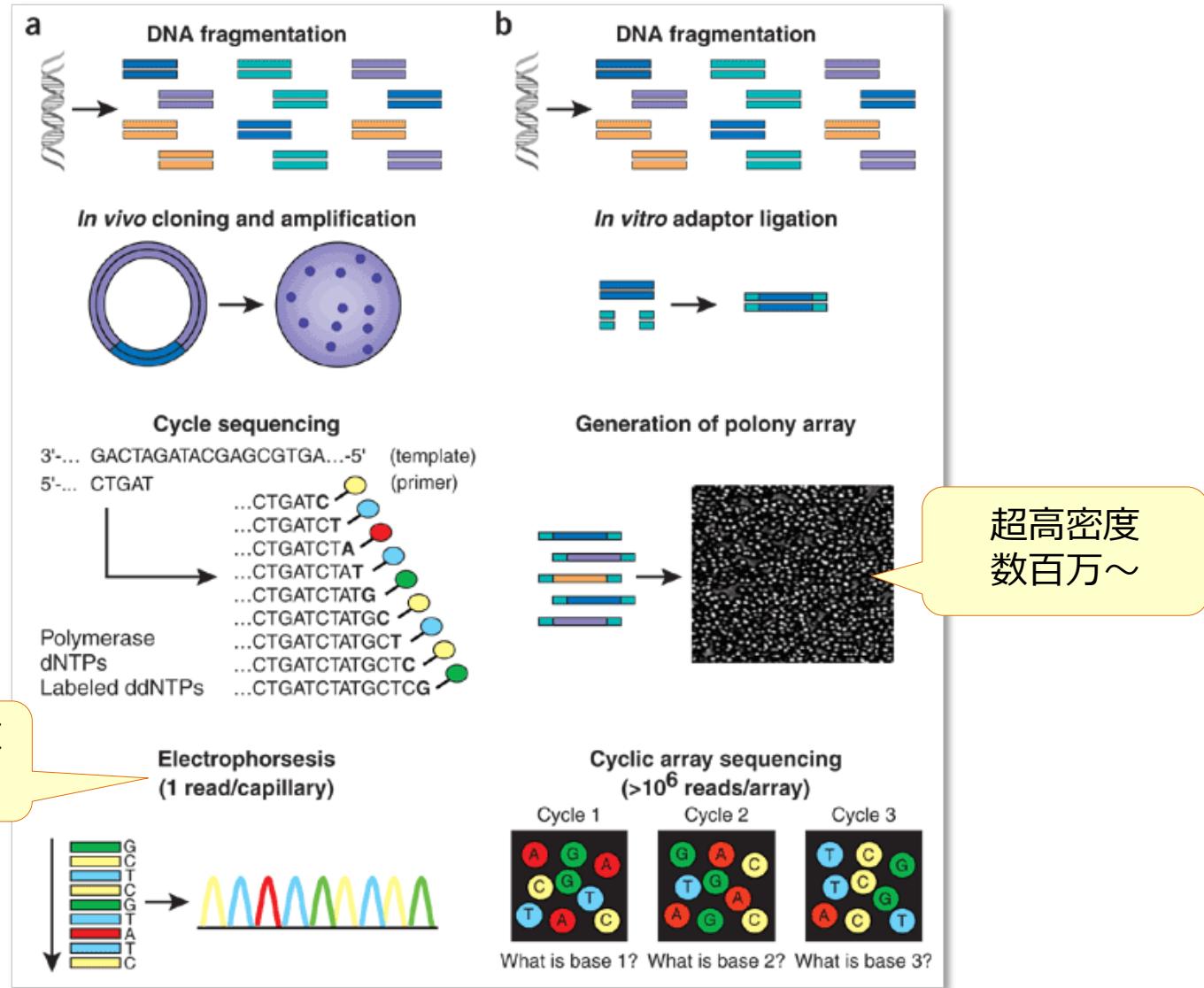
1. 次世代シークエンスデータベースと関連データベース
2. Sequence Read Archive (SRA)
3. EBI/NCBI/DDBJ 各データベースの使い方と特徴
4. アクセス制限データベース
5. 今後の方向性

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)

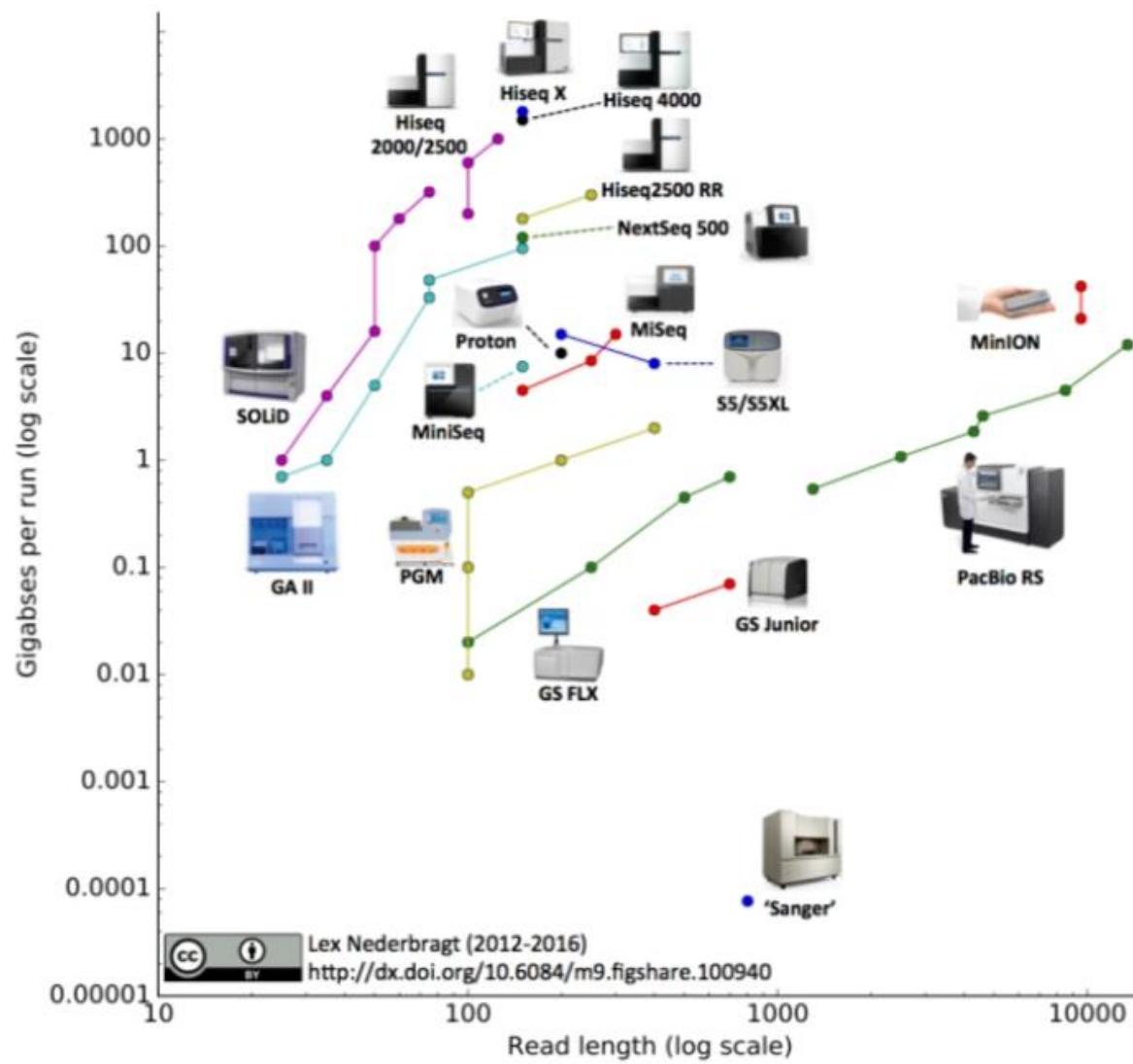


サンガー法 v.s. 次世代シークエンス (NGS)



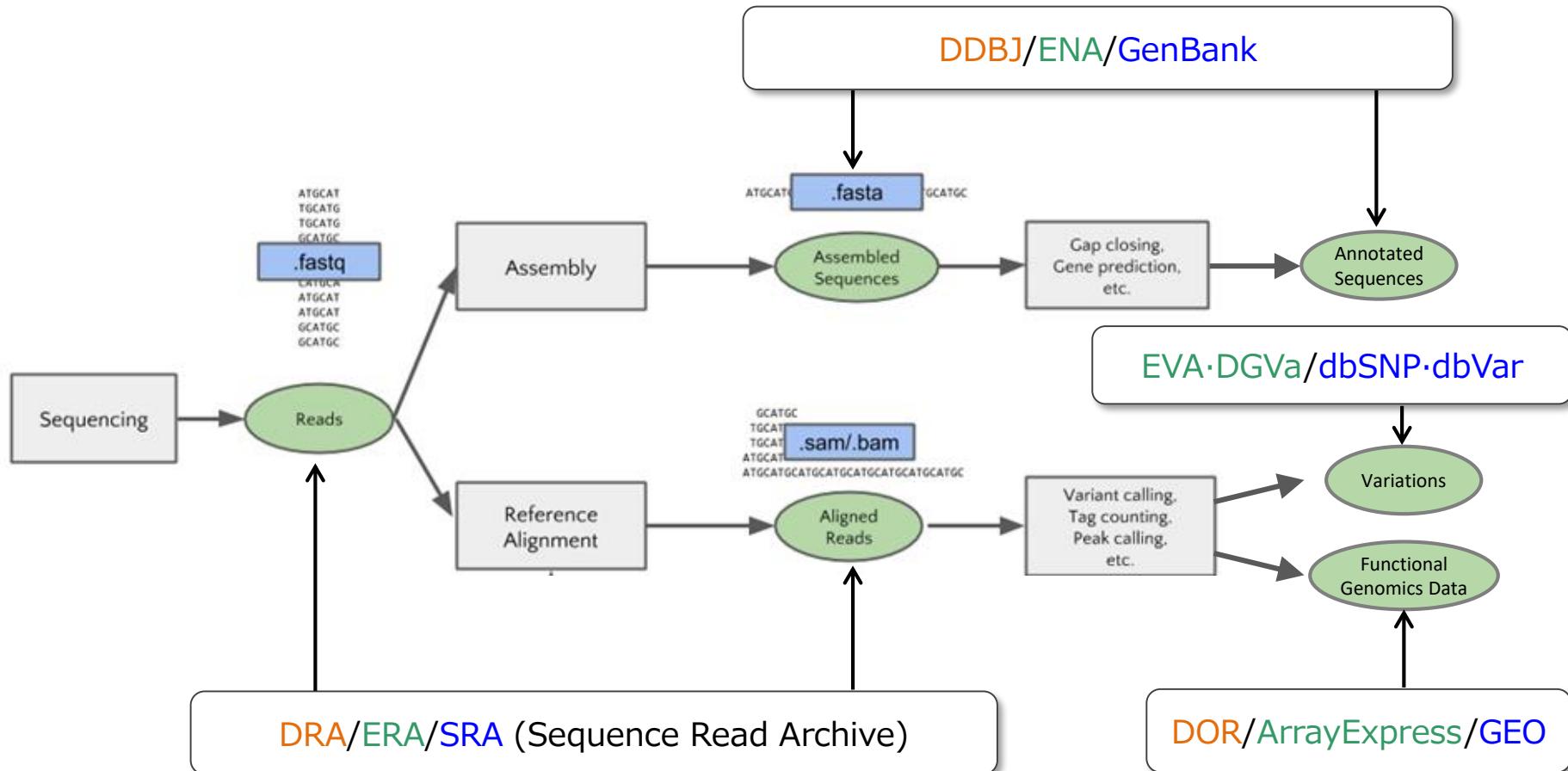
http://www.nature.com/nbt/journal/v26/n10/fig_tab/nbt1486_F1.html

次世代シークエンサー

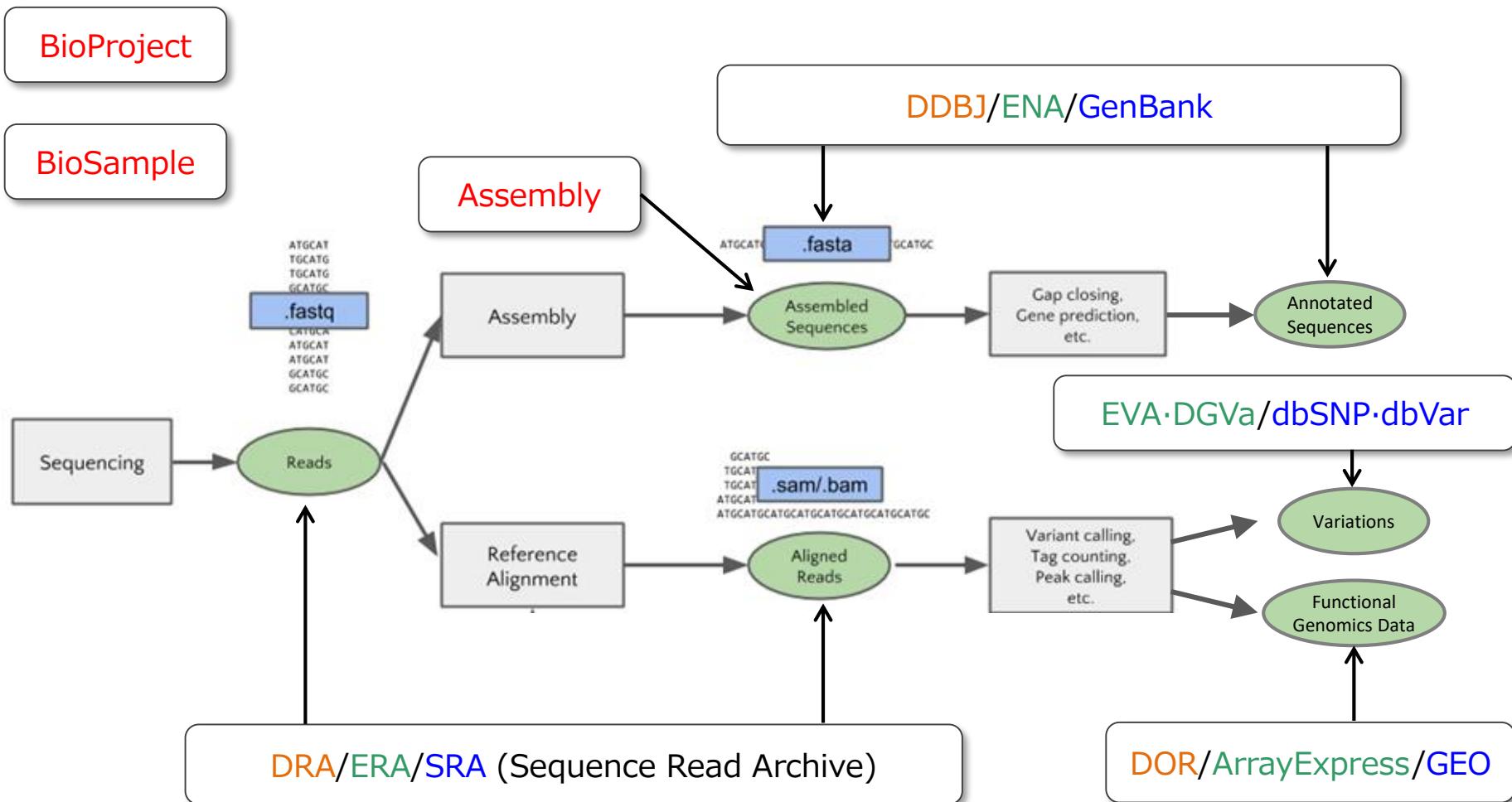


<https://flxlexblog.wordpress.com/2016/07/08/developments-in-high-throughput-sequencing-july-2016-edition/>

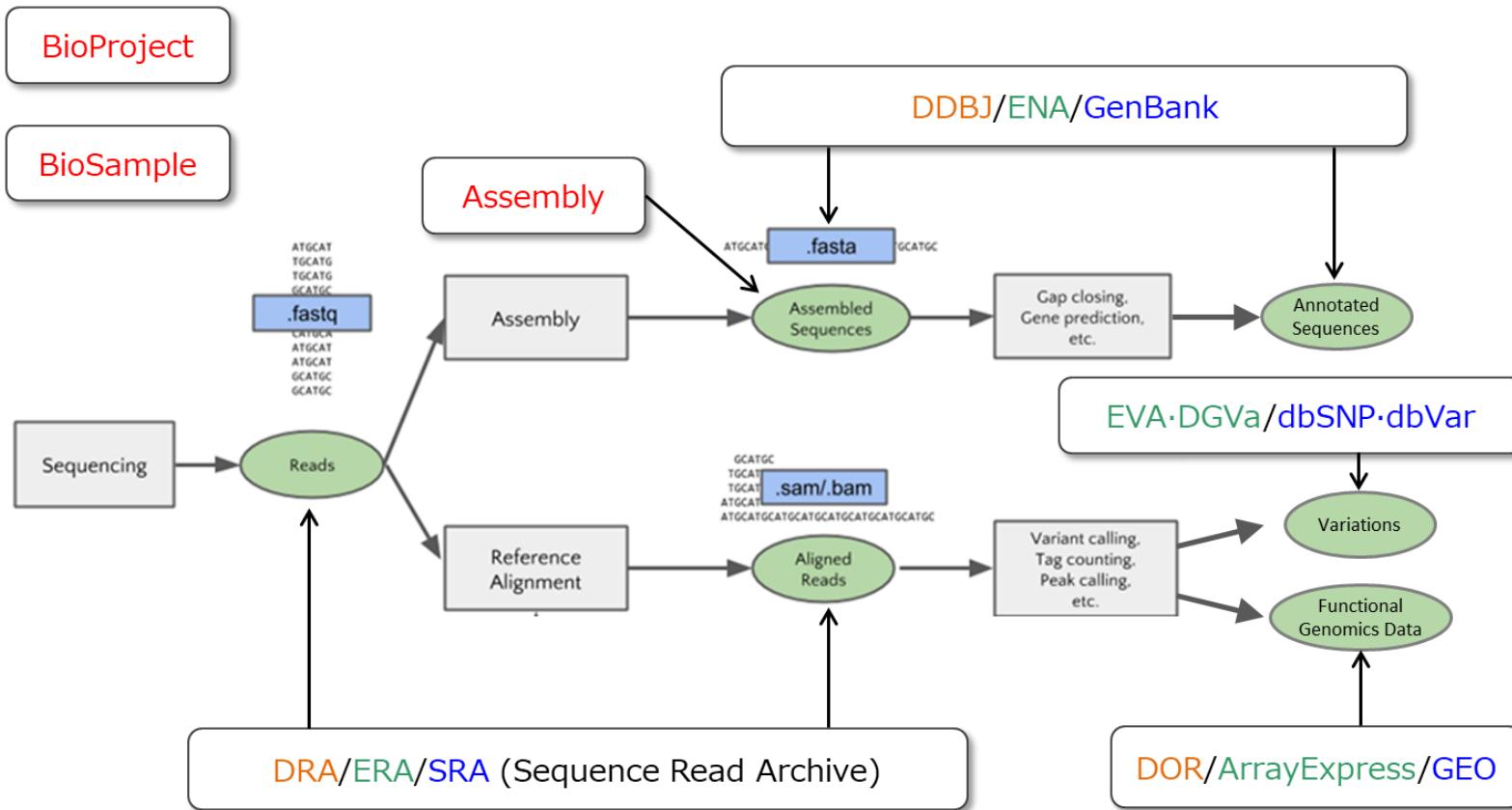
データの種類とデータベース



データを整理するデータベース

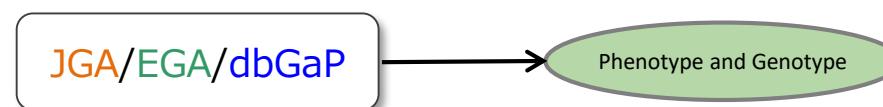


アクセス制限データベース



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• DOR/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/what-database/relational-databases/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	DOR 2018年5月開始		JGA

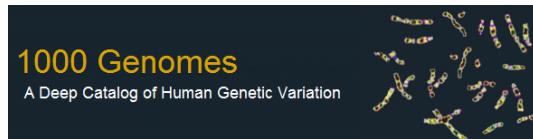
↔ INSDC (国際塩基配列データベース共同事業)



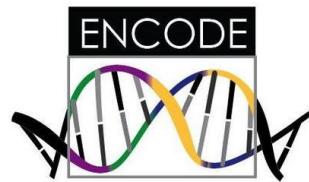
- ↑ データ交換 (アクセス制限はサマリーのみ)
- ↑ データ交換計画中
- ↓ 部分的なデータ取り込み

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

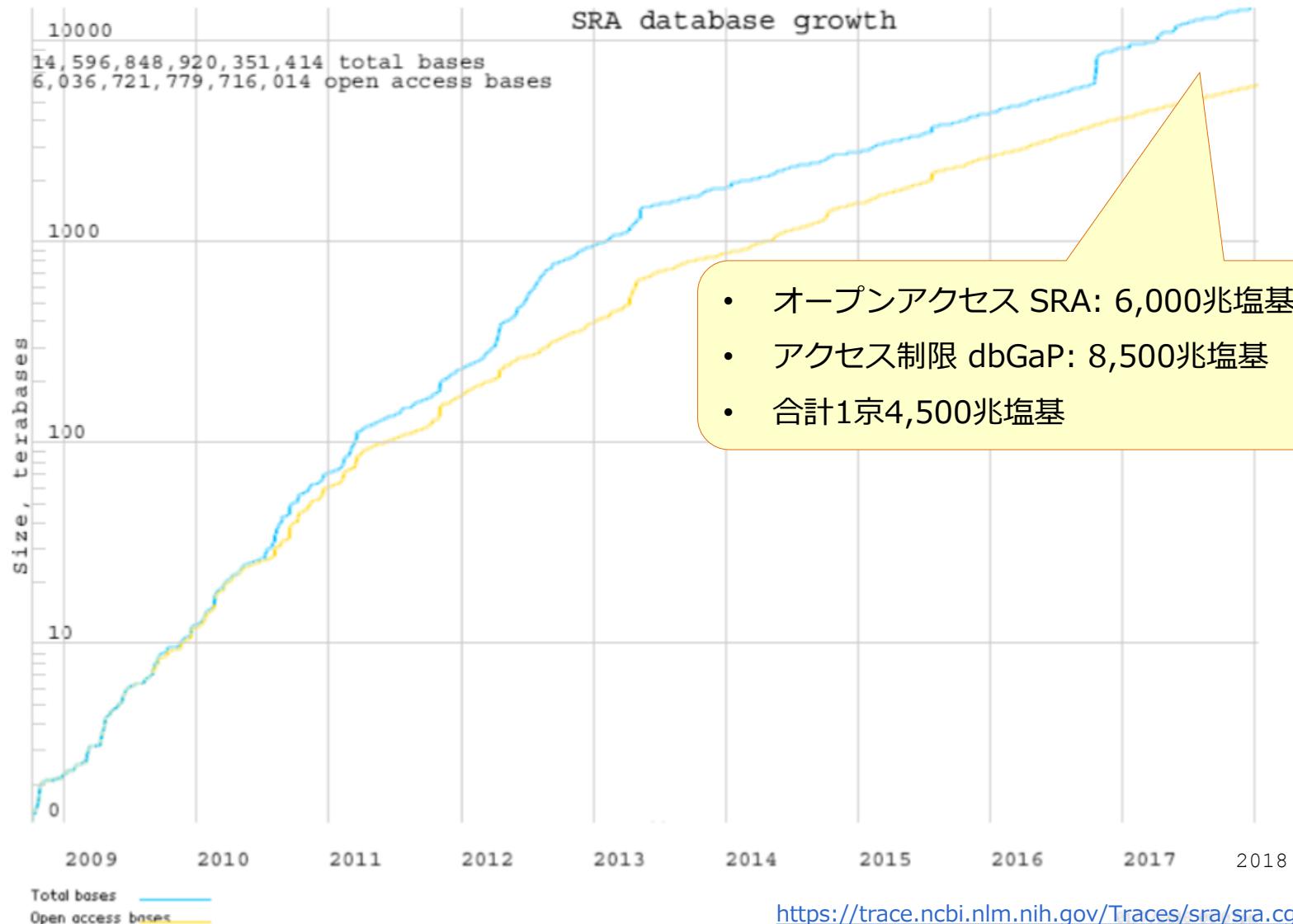


NIH HUMAN
MICROBIOME
PROJECT

- ✓ 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 (DOR/ArrayExpress/GEO)
 - 変異 (EVA·DGVa/dbSNP·dbVar)
 - アクセス制限データベース (JGA/EGA/dbGaP)
- ✓ 共通情報を外部化したデータベース (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)

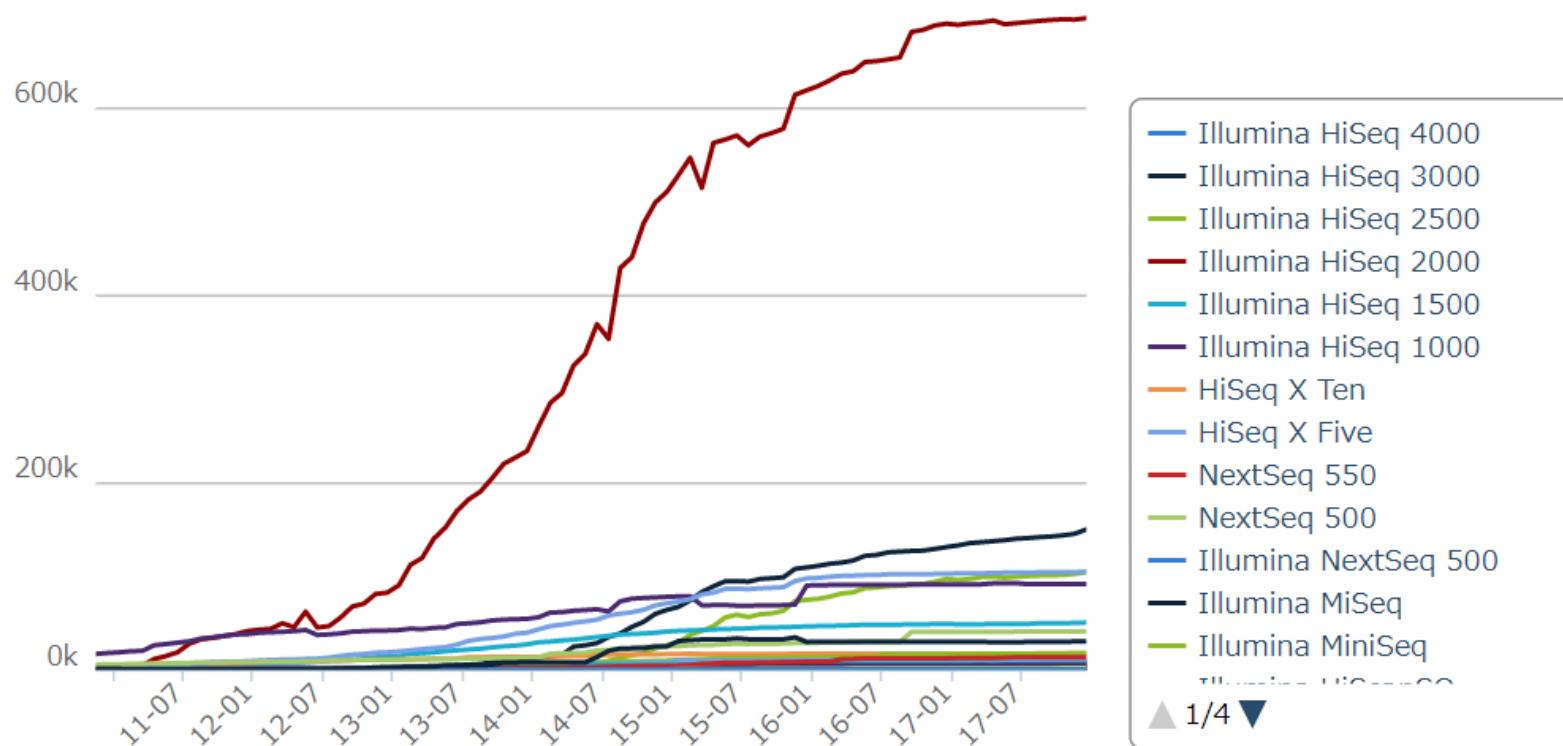
どのくらいのデータがあるか？



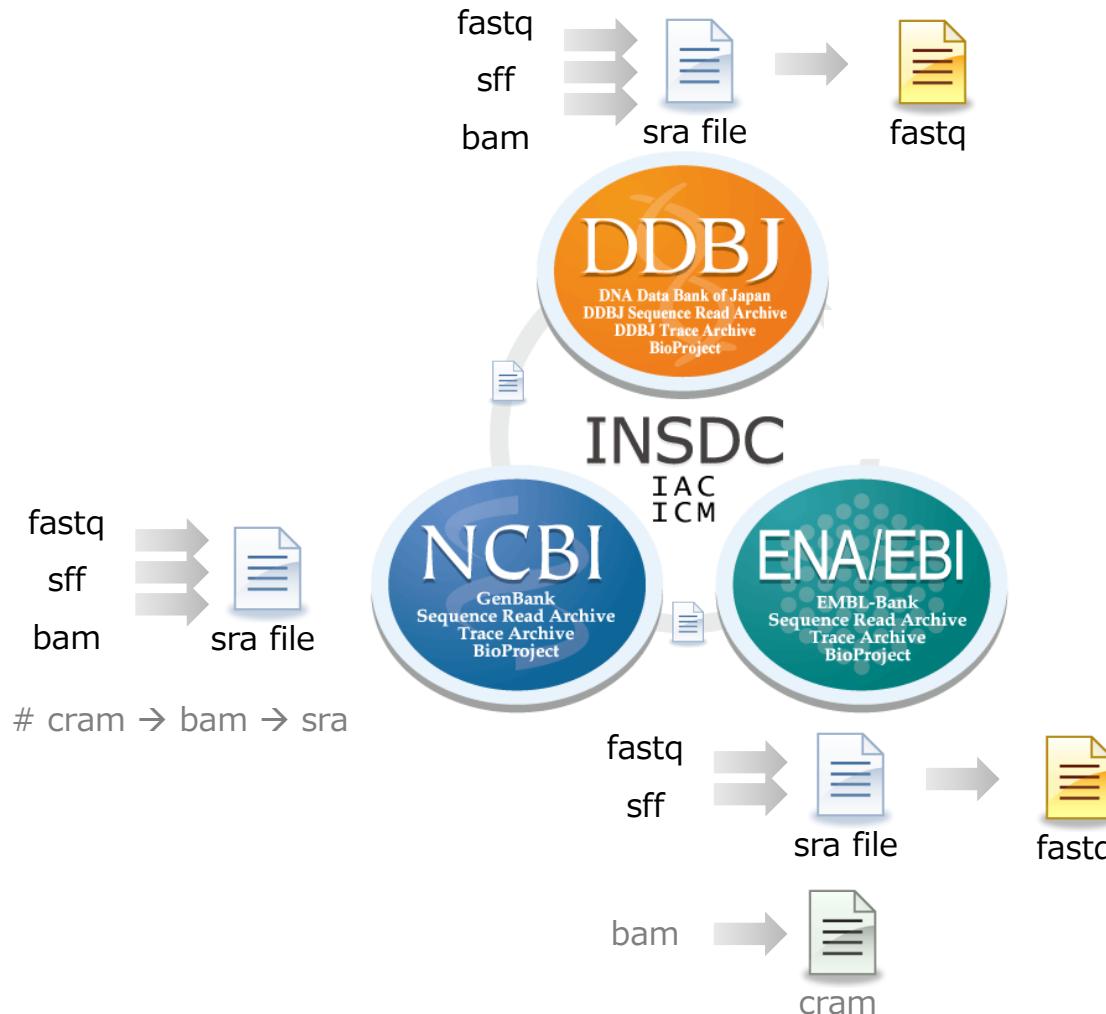
大部分は Illumina

By sequencing platforms

Zoom From To



データファイルフォーマット



- ✓ NCBI が大量リードのコンテナー用に sra ファイルフォーマット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

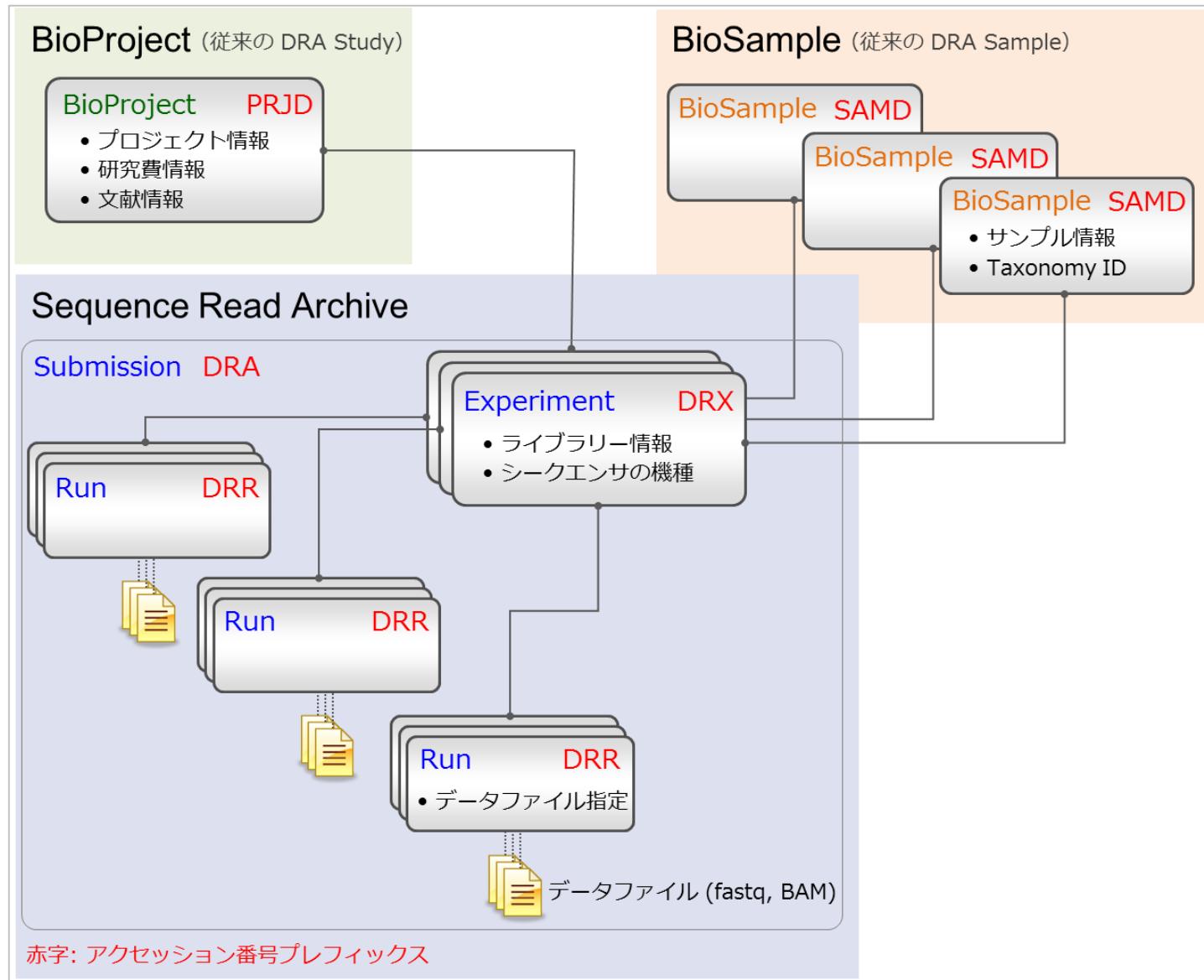
File model	Archive ready?	Stream-able on load?	Auxiliary data?	Run meta data?	Compressed?	Indexed?	Read names parseable ?	Read names indexable?
SRA	Y	Y	Y	Y	Y	Y	Y	Y
BAM	Y	Y	Y	Y	Y	Y	N	N
SFF	Y	Y	Y	Y	Y	Y	Y	Y
HDF5	Y	Y	Y	Y	Y	Y	Y	Y
SOLID	Y	Y	Y	Y	N	N	Y	Y
FASTQ	Y	Y	N	N	N	N	N	N
SRF	N	Y	Y	Y	Y	Y	Y	Y
Illumina native	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	Accession: PRJDB3182 ID: 302624
Data Type	Genome sequencing and assembly	<div style="background-color: #f0f8ff; padding: 5px;"> See Genome Information for <i>Ptychodera flava</i> </div> <div style="background-color: #e0f2fd; padding: 5px;"> NAVIGATE ACROSS 3 additional projects are related by organism. </div>
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:

Download				
Assembly	Level	WGS	BioSample	Taxonomy
GCA_001465055.1	Scaffold	BCFJ000000000	SAMD00023482	<i>Ptychodera flava</i>

SRA Data Details

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

塩基配列

SRA データ

BioSample Assembly

SRA データサイズ

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

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

SRA データ

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

BioSample
Assembly

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																																				
Identifiers	BioSample: SAMD00023482																																				
生物情報	Organism Ptychodera flava cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Hemichordata; Enteropneusta; Ptychoderidae; Ptychodera																																				
パッケージ	Package MIGS: eukaryote; version 4.0																																				
サンプル属性	Attributes <table><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></table> <p>Description Hawaiian Acornworm, Ptychodera flava, Genomic DNA Keywords: GSC:MIxS;MIGS:4.0</p> <p>Links marinegenomicsdb</p> <p>BioProject PRJDB3182 Ptychodera flava Retrieve all samples from this project</p> <p>Submission Okinawa Institute of Science and Technology Marine Genomics Unit; 2015-11-17</p>	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																																				

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; Ptychoderidae; Ptychodera

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)

サンプル履歴 → 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

BioProject →

[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, *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:

シークエンサー →

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

BioProject → 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, *Ptychodera flava*.
[PRJDB3182](#) • [DRP002816](#) • [All experiments](#) • [All runs](#)
[show Abstract](#)

BioSample → Sample: [SAMD00023482](#) • [DRS021946](#) • [All experiments](#) • [All runs](#)

ライブラリー → ✓ ライブラリー + シークエンサー

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)

Run → Spot descriptor:
forward reverse
1 271

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

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

SRA Run

Run 情報 →

リード構成 →

リード →

Illumina MiSeq paired end sequencing of SAMD00023482 (DRR027941)

Metadata Analysis (alpha) Reads Download

Run	Spots	Bases	Size	GC content	Published	Access Type
DRR027941	5.3M	3.0Gbp	1.5G	37.5%	2015-11-17	public

This run has 2 reads per spot:

Legend

Experiment DRX025519 Library Name Platform Strategy Source Selection Layout

to BLAST

Filter: Find Filtered Download What does it do?

What can the filter be applied to?

The Run is too big (>1.1G) for searching by sequence substring.

View: biological reads technical reads quality scores

Reads (separated)

1. DRR027941.1 DRS021946
name: 1, member: 5

2. DRR027941.2 DRS021946
name: 2, member: 5

3. DRR027941.3 DRS021946
name: 3, member: 5

4. DRR027941.4 DRS021946
name: 4, member: 5

5. DRR027941.5 DRS021946
name: 5, member: 5

6. DRR027941.6 DRS021946
name: 6, member: 5

7. DRR027941.7 DRS021946
name: 7, member: 5

8. DRR027941.8 DRS021946
name: 8, member: 5

9. DRR027941.9 DRS021946
name: 9, member: 5

10. DRR027941.10 DRS021946
name: 10, member: 5

>gnl|SRA|DRR027941.4.1 4 (Biological)
NGATAACTTCGTATAATGTATGCTATACGNAGTNATTNNNNNTNGNAAAATTGGAACAGGT
GACAGTATACCTTAAGGCACAGTCAAAATTATCCATGCTAAGATTACGNAGTTATCG

One channel quality score
2 23 33 34 34 38 38 38 38 38 38 38 38 38 38 38 38 38 38 38 38 38 38 38
38
2 25 2 25 34 37 38 39 38 38 38 38 38 38 38 38 38 38 38 38 38 38 38 38
38
38
38 38 38 38 38 38 38 38 38 38 38 38 38 38 38 2 25 34 37 37 37 37 37 38
38 38

>gnl|SRA|DRR027941.4.2 4 (Biological)
NGATAACTTCGTAAATCTTAGCATTGGATAAAATTGACTGTGCCTTAAGGTATACTG
TCACCTGTTCAAATTGCCAACGTAATAACCTCGTATAGCATACATTATACGAAGTTATCG

One channel quality score
2 23 33 34 34 38 38 38 38 38 38 38 38 38 38 38 38 38 38 38 38 38 38 38
38
38
38
38
38 38

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

Run 情報 →

Illumina MiSeq paired end sequencing of SAMD00023482 (DRR027941)

Metadata Analysis (alpha) Reads Download

Run	Spots	Bases	Size	GC content	Published	Access Type
DRR027941	5.3M	3.0Gbp	1.5G	37.5%	2015-11-17	public

This run has 2 reads per spot:

L=281, σ=44.9, 100% L=281, σ=44.9, 100%

Legend

リード構成 →

- ✓ 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)

PRJDB3182

D

Show abstract

リード →

1. DRR027941.1 DRS021946 name: 1, member: 5
2. DRR027941.2 DRS021946 name: 2, member: 5
3. DRR027941.3 DRS021946 name: 3, member: 5
4. DRR027941.4 DRS021946 name: 4, member: 5
5. DRR027941.5 DRS021946 name: 5, member: 5
6. DRR027941.6 DRS021946 name: 6, member: 5
7. DRR027941.7 DRS021946 name: 7, member: 5
8. DRR027941.8 DRS021946 name: 8, member: 5
9. DRR027941.9 DRS021946 name: 9, member: 5
10. DRR027941.10 DRS021946 name: 10, member: 5

>gnl|SRA|DRR027941.4.1 4 (Biological)

NGATAACTTCGTATAATGTATGCTATACGNAGTNATTNNNNNTNGAAAAATTGGAACAGGT
GACAGTATACCTTAAGGCACAGTCAAAATTATTCCAATGCTAAGATTACGNAGTTATCG

One channel quality score

2 23 33 34 34 38 38 38 38 38 38 38 38 38 38 38 38 38 38 38 38 38
38 38 38 38 38 38 38 38 2 28 35 37 2 25 34 37 2 2 2
2 25 2 25 2 25 34 37 38 39 38 38 38 38 38 38 38 38 38 38 38 38
38
38
38
38
38 38

>gnl|SRA|DRR027941.4.2 4 (Biological)

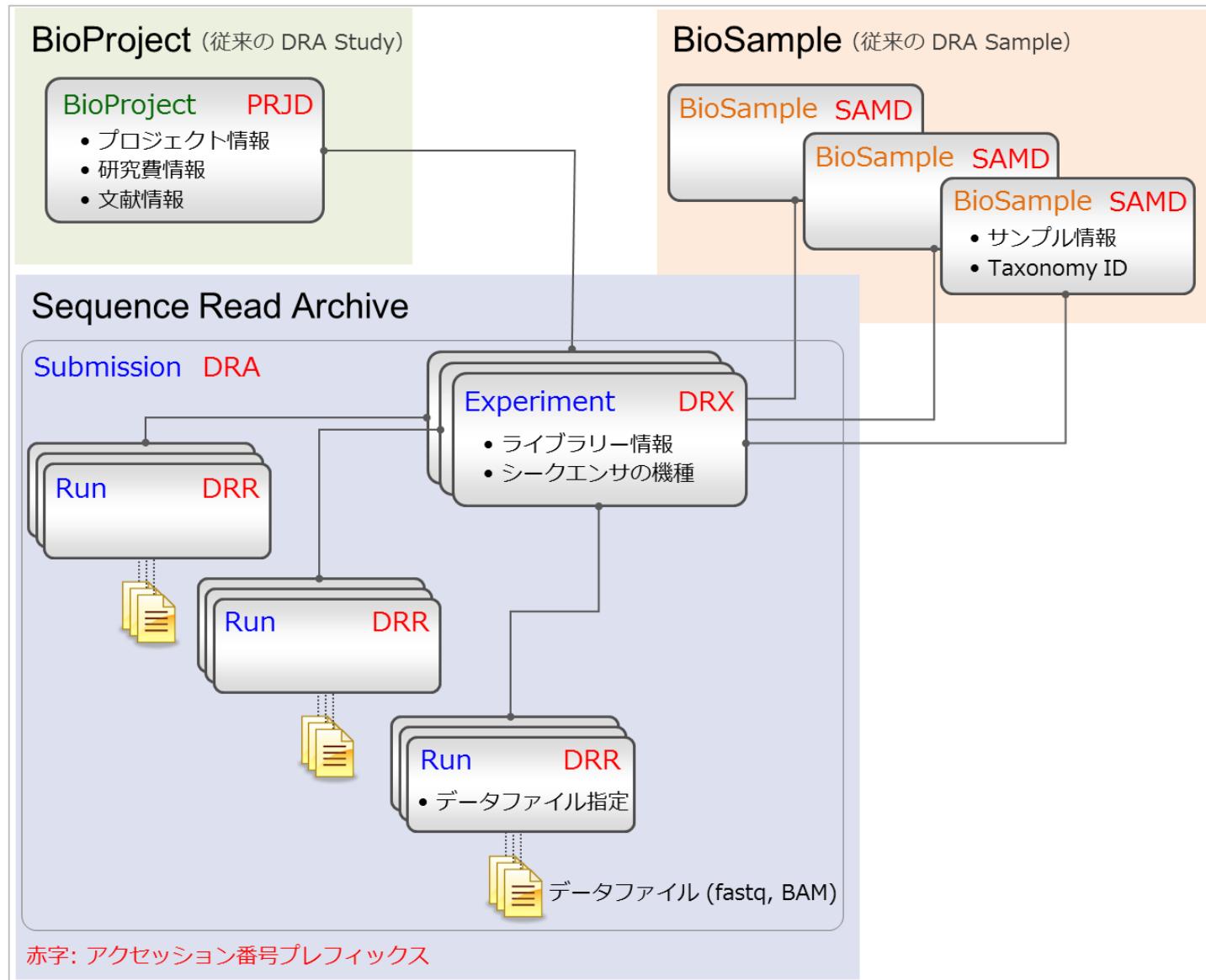
NGATAACTTCGTAAATCTTAGCATTGGATAAAATTGACTGTGCCTTAAGGTATACTG
TCACCTGTTCAATTTCGCCAACGTAATAACCTTCGTATAGCATACATTACGAAGTTATCG

One channel quality score

2 23 33 34 34 38 38 38 38 38 38 38 38 38 38 38 38 38 38 38 38
38
38
38
38
38
38 38

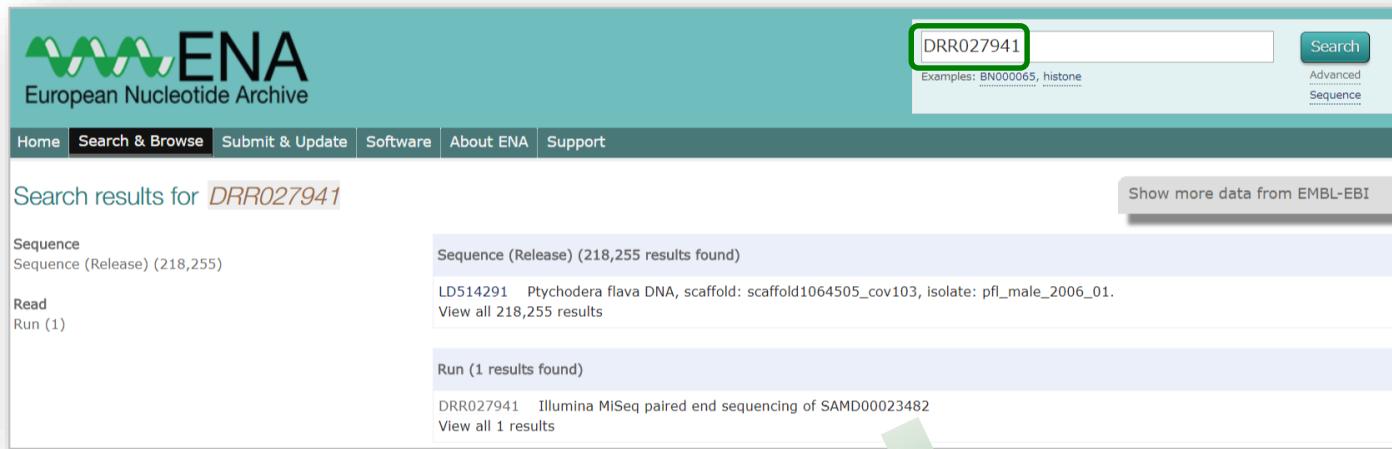
NCBI Run: <https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=DRR027941>

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)



Search results for *DRR027941*

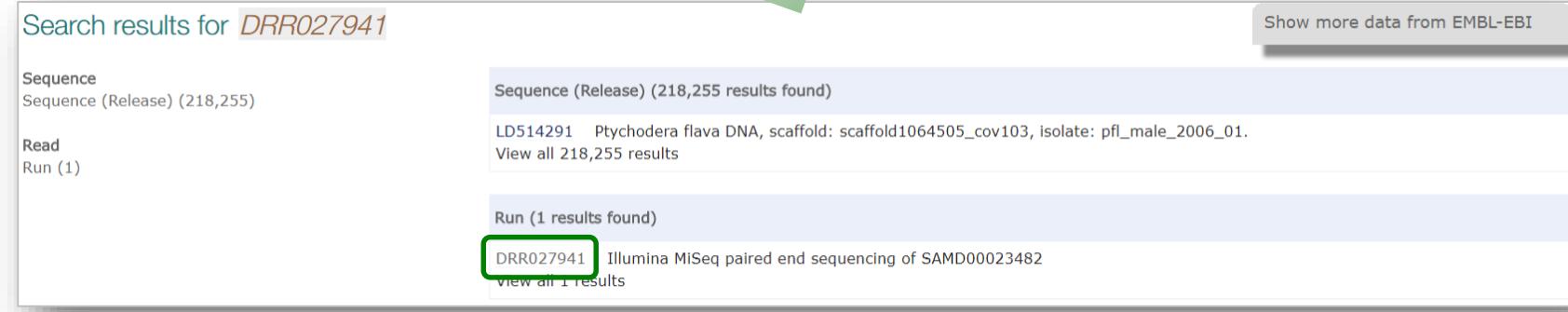
Sequence
Sequence (Release) (218,255)

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 another browser. 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"/> Submitted bytes <input type="checkbox"/> Submitted bytes <input type="checkbox"/> Submitted files (Aspera) <input checked="" type="checkbox"/> Submitted files (Galaxy) <input type="checkbox"/> Submitted files (FTP) <input checked="" type="checkbox"/> NCBI SRA file (FTP) <input type="checkbox"/> Submitted MD5 <input type="checkbox"/> CRAM Index files (Aspera) <input type="checkbox"/> Submitted format <input type="checkbox"/> Sample title <input type="checkbox"/> Submitter's sample name <input type="checkbox"/> Sample title <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 is 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)	NCBI SRA files (Galaxy)
PRJDB3182	SAMD00023482	DRS021946	DRX025509	DRR027930	63121	Ptychodera flava	454 GS FLX	SINGLE	File 1	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	File 1
PRJDB3182	SAMD00023482	DRS021946	DRX025511	DRR027932	63121	Ptychodera flava	454 GS FLX+	SINGLE	File 1	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	File 1
PRJDB3182	SAMD00023482	DRS021946	DRX025513	DRR027934	63121	Ptychodera flava	Illumina HiSeq 2000	PAIRED	File 1	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	File 1
PRJDB3182	SAMD00023482	DRS021946	DRX025515	DRR027936	63121	Ptychodera flava	Illumina MiSeq	PAIRED	File 1	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	File 1
PRJDB3182	SAMD00023482	DRS021946	DRX025517	DRR027938	63121	Ptychodera flava	Illumina MiSeq	PAIRED	File 1	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	File 1

Prev Next Go to page: 1 Go

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

Study 配下の全オブジェクト

ENAFTP Downloader

Local Download Folder: C:\ena Browse

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) Sequence Read Archive Submission (DRA) Experiment (DRX) Run (DRR) データファイル配下

ナレーターの説明：「Study 配下の全オブジェクト」と「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」が、それぞれ図中の緑色の丸と青色の丸で示されています。また、右側の「ENAFTP Downloader」の操作画面には、複数のデータファイル（FASTQ）がリストされ、「4 FASTQ files selected. Total size: 11」と表示されています。

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

✓ 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 are two buttons: "Search" and "Advanced Sequence". The "Advanced Sequence" button is circled in green, and a large green arrow points down to the search form below.

Popular

Advanced Upload accession

Search query [Help](#)

Text Search

Examples: BN000065, histone

[Advanced search](#)

Sequence Search

Enter or paste a nucleotide sequence or accession number

[Advanced search](#)

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

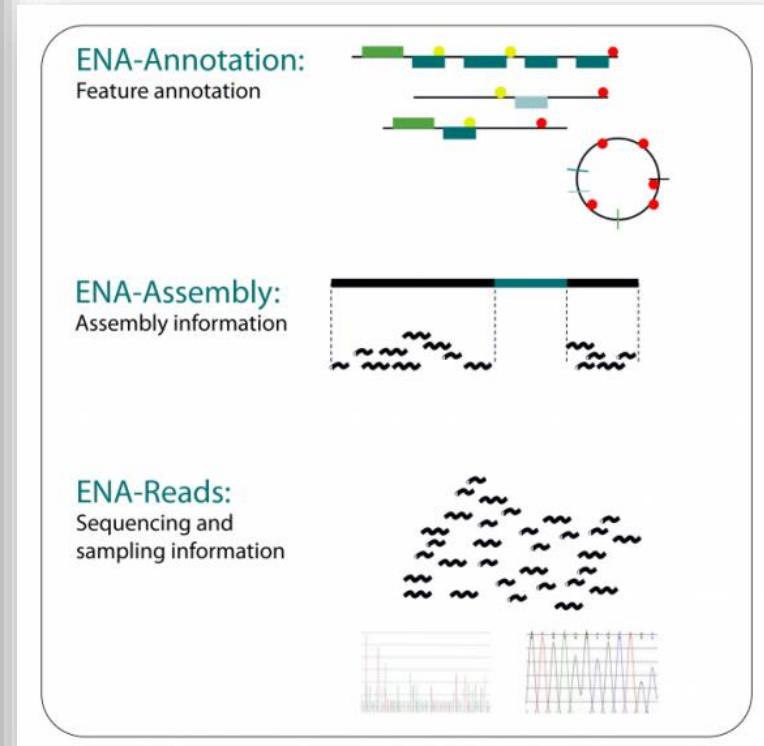
Advanced Upload accession

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:

- 2018 Jan
- 2018 Feb

✓ 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	39946	Oryza sativa Indica Group	PAIRED
					9	4530	Oryza sativa	SINGLE
					20	39947	Oryza sativa	PAIRED
					19767	39947	Oryza sativa Japonica Group	SINGLE
					19768	39947	Oryza sativa	PAIRED
					19769	39947	Oryza sativa Japonica Group	SINGLE
					19770	39947	Oryza sativa	PAIRED
					19771	39947	Oryza sativa Japonica Group	SINGLE
					19772	39947	Oryza sativa	PAIRED
					19773	39947	Oryza sativa Japonica Group	SINGLE
					19774	39947	Oryza sativa	PAIRED
					19775	39947	Oryza sativa Japonica Group	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: 1 - 1 of 1 results in TEXT

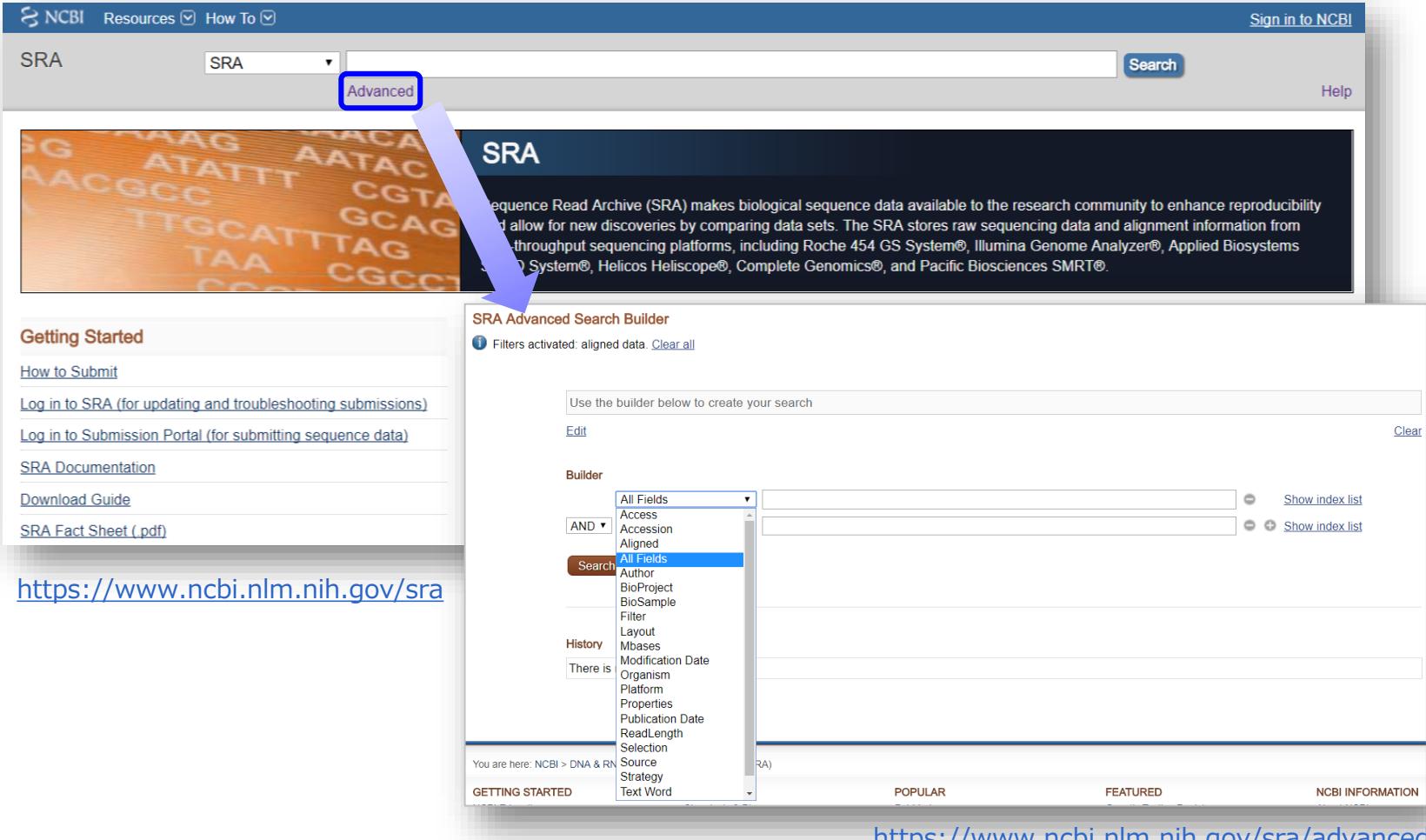
Select columns

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)
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 main search area has a dropdown menu set to "SRA" and a search input field with the word "SRA". A blue box highlights the "Advanced" button above the search input field. A large blue arrow points from this "Advanced" button down to the "SRA Advanced Search Builder" panel.

SRA Advanced Search Builder

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

Use the builder below to create your search

Edit [Clear](#)

Builder

All Fields AND All Fields

Search History There is

- All Fields
- Access
- Accession
- Aligned
- Author
- BioProject
- BioSample
- Filter
- Layout
- Mbases
- Modification Date
- Organism
- Platform
- Properties
- Publication Date
- ReadLength
- Selection
- Source
- Strategy
- Text Word

You are here: NCBI > DNA & RNA > SRA

POPULAR FEATURED NCBI INFORMATION

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

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

✓ 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 the query "Homo sapiens". The interface includes a sidebar with filters for Access (Controlled vs Public), Source (DNA/RNA), Type (exome/genome), and Other (aligned data). The main search results list 12 items, each detailing a metagenomic run from LS454 (454 GS FLX) with various statistics like number of spots, bases, and downloads. To the right, three sections are highlighted with blue arrows: "Taxonomy ランキング" pointing to a tree view of organisms; "BioProject ランキング" pointing to a list of top BioProjects; and "関連データベース レコード" pointing to a section of related database records.

Access
Controlled (288,116)
Public (911,075)

Source
DNA (939,706)
RNA (259,510)

Type
exome (211,169)
genome (351,823)

Other
aligned data (383,851)

Search fields
Choose ...

Clear all

Show additional filters

SRA SRA ▾ Homo sapiens Create alert Advanced Search Sign in to NCBI Help

Items: 1 to 20 of 1211547

Send to: Filters: Manage Filters

Results by taxon

Top Organisms [Tree]

- Homo sapiens (925082)
- human gut metagenome (68301)
- Salmonella enterica (28637)
- Mycobacterium tuberculosis (17382)
- human metagenome (15457)
- All other taxa (156688)

More...

Top Bioprojects

- NIH Epigenomics Roadmap Initiative (1893)
- Production ENCODE epigenomic... (1493)
- Production ENCODE functional... (1137)
- Production ENCODE transcript... (363)

Search in related databases

Database	Access	public	controlled	all
BioSample	690,559	286,916	977,475	
BioProject	10,965	483	11,448	
dbGaP		1	1	
GEO Datasets	212,631		212,631	

Find related data

Database: Select

Find items

Search details

"Homo sapiens" [Organism] OR Homo sapiens [All Fields]

Search See more...

Recent activity

Turn Off Clear

Homo sapiens (1211547)

human (1293205)

See more...

<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 <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	• 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

1. [Illumina MiSeq paired end sequencing of SAMD00023482](#)
Accession: DRX025519

2. [Illumina MiSeq paired end sequencing of SAMD00023482](#)
Accession: DRX025518

3. [Illumina MiSeq paired end sequencing of SAMD00023482](#)
Accession: DRX025517

4. [Illumina MiSeq paired end sequencing of SAMD00023482](#)
Accession: DRX025516

5. [Illumina MiSeq paired end sequencing of SAMD00023482](#)
Accession: DRX025515

6. [Illumina MiSeq paired end sequencing of SAMD00023482](#)
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: *Ptychodera flava*
- ReleaseDate: 2015-11-17
- SRA Study: DRP002816
- Sample Name: DR5021946
- bioProject Id: PRJDB3182
- collection date: 2006-12-10
- description: Hawaiian Acornworm, *Ptychodera 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: *Ptychodera flava* Genomic DNA
- propagation: NA
- ref biomaterial: NA
- sample name: *Ptychodera flava* Genomic DNA
- sample title: *Ptychodera flava* Genomic DNA

Run Information Table (bottom left):

Runs	Bytes	Bases	Download
Total: 27	212.44 Gb	273.37 G	RunInfo Table Accession List

Run Details 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

Annotations:

- Facets 検索 (Facets)
- 共通情報 (Common Information)
- Run 情報の表をファイルでダウンロード (Run Information Table download)
- アクセスション番号リストをファイルでダウンロード (Accession List download)
- Run と参照オブジェクトの表 (Run and Reference Object Table)

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

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

染色体を選択

この Run を Sequence Viewer で表示

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

Accession: PRJNA59845 ID: 59845

See Genome Information for Homo sapiens

NAVIGATE UP

This project is a component of the 1000 Genomes Full Production Exome Sequencing

NAVIGATE ACROSS

36 additional projects are components of the 1000 Genomes Full Production Exome

Summary ▾

Send results to Blast

Links from BioProject

Items: 3

Filters activated: aligned data. Clear all

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

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

Illumina random exon sequencing of genomic DNA paired-end library 'Pond-314377' containing sample 'NA18939'

2 ILLUMINA (Illumina HiSeq 2000) runs: 24.8M spots, 3.8G bases, 1.7Gb downloads

Accession: SRX655425

Metadata Alignment Analysis (alpha) Reads Download

Alignment Reads Bases Fraction

Primary 6.3M 478.7Mbp 98.64%

Reference

Homo sapiens chromosome 12, GRCh37.p13 Primary Assembly

View scope accession count in Sequence Viewer

this run SRR1517928 1

same experiment SRX655465 1

same sample SRS000715

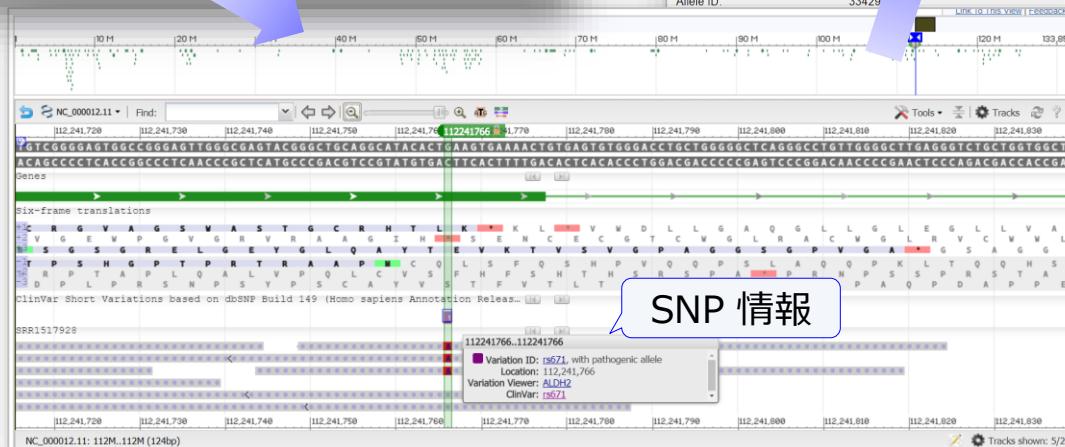
same study SRP004076

all sra

Output this run in FASTA format to Screen Fasta

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

NCBI: Sequence viewer



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

ClinVar: 変異の医学的な解釈
ALDH2, E504K (rs671)
GG: 高活性 日本人の56%
GA: 低活性 40%
AA: 非活性 4%

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

- ✓ 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: 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 identifying the sample in the SRA interface, then navigating to the BLAST search page, and finally executing the search.

Left Panel (SRA Interface):

- Run Details:** DRR068403, 7.3k Spots, 2.6Mbp, 1.6M Size, 52.9% GC content, Published 2016-11-17, Access Type public.
- Experiment Table:** DRX062354, Library Name, Platform Illumina, Strategy AMPLICON, Source GENOMIC, Selection PCR, Layout SING. A blue box highlights the "to BLAST" button.
- Biosample:** SAMD00058608 (D), Human gut metagenome.
- Bioproject:** PRJDB4998, DRP003338, Fecal microbiota of Japanese healthy adults.
- Sample Details:** Microbiota of the fecal sample from subject 516. Includes identifiers, organism (human gut metagenome), package (MIMS), 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), and submission (Yakult Central Institute; 2016-11-17).

Middle Panel (BLAST Search Page):

- Search Set:** SRA Experiment set (SRX) (Job title: Bifidobacterium 16S ribosomal RNA).
- Program Selection:** BLASTN.
- BLAST Button:** The "BLAST" button is highlighted with a blue arrow pointing from the SRA interface.

Right Panel (BLAST Results):

- Graphic Summary:** Distribution of the top 100 Blast Hits on 406 subject sequences. A color key indicates alignment scores: <40 (black), 40-50 (blue), 50-80 (green), 80-200 (magenta), >=200 (red). The plot shows a dense vertical column of red bars, indicating high similarity across most sequences.
- Descriptions:** Sequences producing significant alignments. A table lists three entries: DR0002354, DR0002354, and DR0002354. Columns include 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

Illumina MiSeq sequencing of SAMD00058608 (DRR068403)

[Metadata](#) **Analysis (alpha)** [Reads](#) [Download](#)

Run	Spots	Bases	Size	GC content	Published	Access Type
DRR068403	7.3k	2.6Mbp	1.6M	52.9%	2016-11-17	public

This run has 2 reads per spot:

L=353, 100% L=0

Legend

Experiment	Library Name	Platform	Strategy	Source	Selection	Layout
DRX062354		Illumina	AMPLICON	GENOMIC	PCR	SINGLE
to BLAST						

Biosample Sample Description Organism

SAMD00058608 (DRS035752) human gut metagenome

Bioproject SRA Study Title

PRJDB4998 DRP003338 Fecal microbiota of Japanese healthy adults

Abstract:

Fecal microbiota of Japanese healthy adults. 16s rRNA gene library of V1-V2 hyper variable regions derived from the fecal samples of 516 subjects.

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

Metadata Analysis (alpha) Reads Download

Warning: experimental software

Taxonomy Analysis

- Unidentified reads: **4.34%**
- Identified reads: **95.66%**
 - Bacteria: **95.66%**
 - FCB group: **39.11%**
 - Bacteroidales: **39.11%**
 - Bacteroidaceae: **30.05%**
 - Bacteroides: **30.05%**
 - Bacteroides thetaiotaomicron: **0.06%**
 - Rikenellaceae: **2.43%**
 - Prevotellaceae: **0.06%**
 - Terrabacteria group: **17.07%**
 - Firmicutes: **10.49%**
 - Clostridia: **9.81%**
 - Clostridiales: **6.27%**
 - Lachnospiraceae: **3.9%**
 - Anaerostipes: **0.11%**
 - Erysipelotrichia: **0.04%**
 - Eggerthia catenaformis OT 569 = DSM 20559: **0.04%**
 - Actinobacteria: **4.73%**
 - Actinobacteria: **4.67%**
 - Bifidobacterium: **4.66%**
 - Coriobacteria: **0.04%**
 - Collinsella: **0.04%**
 - Proteobacteria: **1.02%**
 - delta/epsilon subdivisions: **0.07%**
 - Desulfovibrionaceae: **0.07%**
 - Gammaproteobacteria: **0.06%**
 - Salmonella enterica: **0.06%**

ビフィズス菌

Strong signals

SuperKingdom	Organism	Rank	% Kbp	weighted score
Bacteria	Bacteroidales order	77.6	1,986	2.0

How to read results?
How taxonomy analysis is done?

Assembly

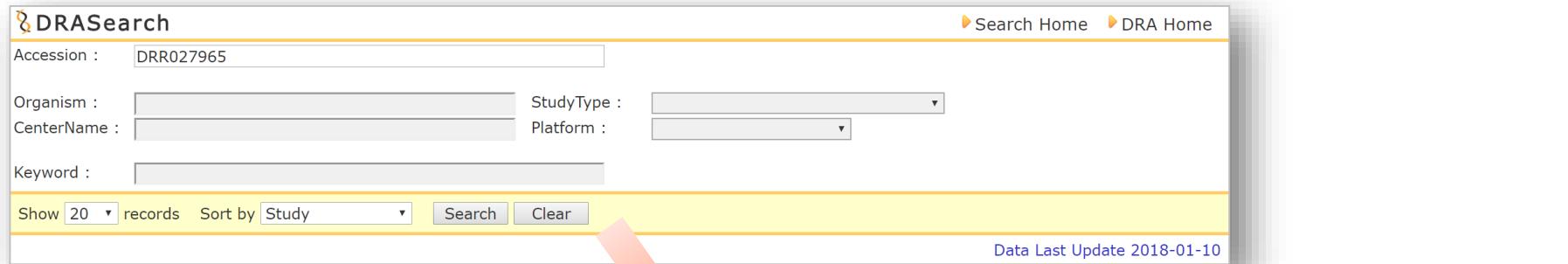
You can also check contigs assembled from kmers of this run:

 - contigs: [DRR068403.contigs.fasta](#) (39 Kb)

How to read contig names?
How it's built?

✓ リードの taxonomy クラスタリング結果（アルファ版）

DDBJ Sequence Read Archive



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

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 URL above to the 'FASTQ' and 'SRA' download links in the main content area. The content area displays 'Run Detail' for DRR027965, including Alias, Instrument model, Date of run, Run center, Number of spots (4,653), and Number of bases (1,190,554). It also shows a sequence viewer with a 'READS (joined)' section and a 'quality' section. To the right is a 'Navigation' sidebar with links for Submission (DRA002856), Study (DRP003382), Experiment (DRX025528), and Sample (DRS037503). A second red arrow points from the 'FTP' link in the navigation sidebar to a detailed file list for experiment DRX025528. This list includes files like DRA002856.experiment.xml, DRA002856.run.xml, DRA002856.sample.xml, DRA002856.study.xml, DRA002856.submission.xml, DRX025520/, DRX025521/, DRX025522/, DRX025523/, DRX025524/, and DRX025525/. The last file listed is DRX025525/ with a timestamp of 2017/04/14 9:00:00.



✓ 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

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

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



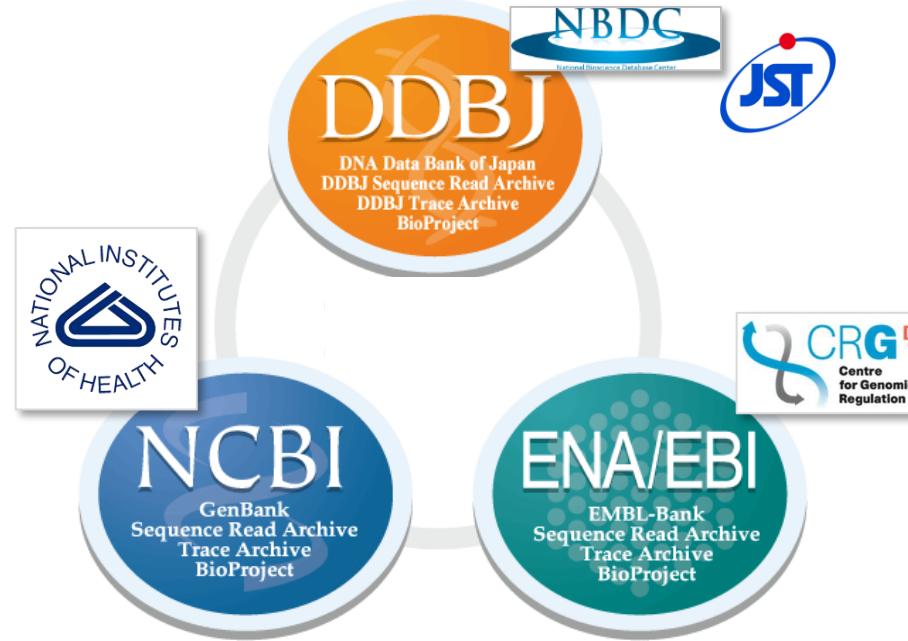
- ✓ NCBI 各サービスとの高度なインテグレーション、特に ClinVar 等の医学系リソース
- ✓ アライメントデータは viewer でブラウズできる
- ✓ sra ファイルから fastq etc への変換が必要



- ✓ 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
- ✓ 他のサービスとの連携はこれから

4. アクセス制限データベース

JGA Japanese Genotype-phenotype Archive



dbGaP

Database of Genotype and Phenotype

EGA

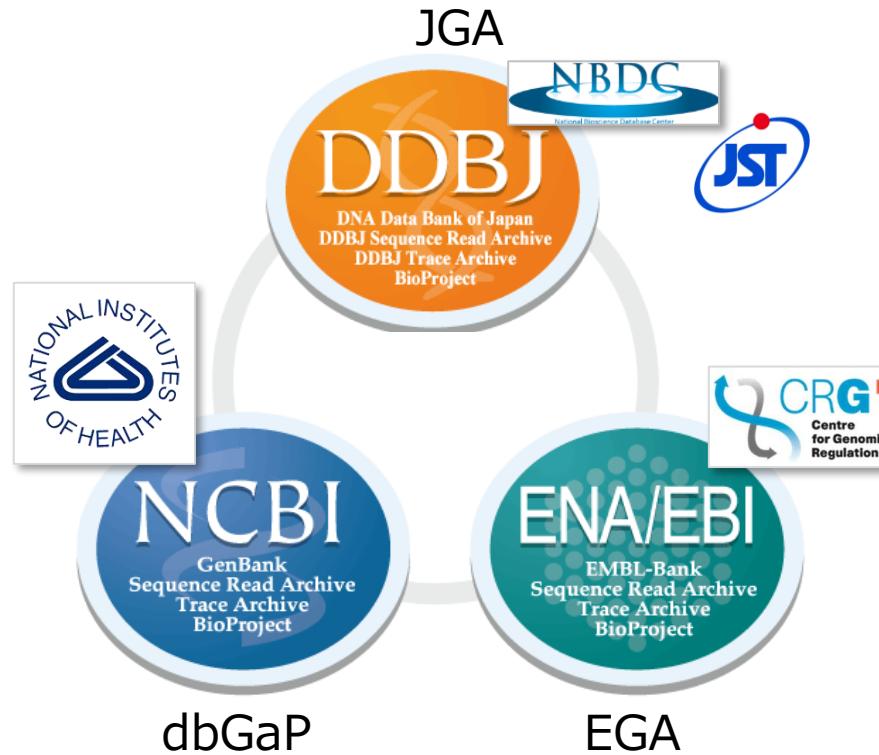
European Genome-phenome Archive

- ✓ JGA と EGA は SRA をベースにした同様のデータモデルを使用
- ✓ dbGaP と EGA はサマリー情報を交換

JGA も参加予定

データの利用審査体制

- ✓ NBDC が Data Access Committee (DAC)



- ✓ NIH の各 Institutes が DAC

- ✓ データ提供組織やファンド機関が DAC

NCBI Resources How To Sign in to NCBI

dbGaP dbGaP Search Help

Limits Advanced

dbGaP
The database of Genotype and Phenotype from studies that have been made public.

Framingham Cohort
dbGaP Study Accession: phs000007.v29.p10

Show BioProject List
Study Variables Documents Analyses Datasets Molecular Data

Variable Name and Accession
Variable Name: STUDY
Variable Accession: phv00159482.v15.p10
Variable belongs to dataset: phd01415.v17.phd11: Framingham, Sample: SHaRe subject ID to sample mapping. This subject to sample mapping includes samples from all of the Framingham substudies, i.e., SHaRe, GAGE, SABRe, Medical Resequencing, ESP, HeartGO, CHARGE-S, and DNA Methylation. This table also includes DNA Methylation substudy and Coriell HapMap samples that were used as substudy controls. Additionally, there is a mapping of sample IDs to other sample ID aliases, the substudy [phs] accession that the sample belongs to, and sample use.

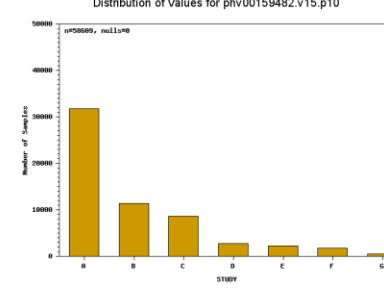
Variable version history

Variable Description
DbGaP top-level study or substudy accession

Statistical Summary
View Summary by Consent Group: Total
 Health/Medical/Biomedical (RB, MDS)
 Health/Medical/Biomedical (RB, NP, MDS)

Distribution of Values for phv00159482.v15.p10

Number of Samples: n=58669, null<=8



STUDY	Number of Samples
A	~35,000
B	~12,000
C	~8,000
D	~4,000
E	~3,000
F	~2,000
G	~1,000

Variables
Search Within This Study
Variables
Framingham Cohort
Validated / Reviewed / Scored / Abstracted Data
Tests
Lab Work
Clinic Questionnaire (Interview and Physical Exam)
Administration

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

https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/variable.cgi?study_id=phs000007.v29.p10&phy=159482&phd=1105&pha=4313&pht=1415&phvf=&phdf=&phaf=&phtf=&dssp=1&consent=&temp=1

- ✓ 970 studies, 1,637,936 subjects (2018年1月11日)
- ✓ よくキュレーションされたリッチなデータ、NCBI 他リソースとの連携

NCBI dbGaP: Advanced Search

The screenshot shows the NCBI dbGaP website. On the left, there's a sidebar with links like 'Access dbGaP Data', 'Resources', and 'Advanced Search'. The main area is titled 'dbGaP Advanced Search' and contains several filter sections: 'Study Disease/Focus', 'Study Design', 'Study Molecular Data Type', 'Study Markerset', 'NIH Institute', 'Study Consent', 'Study Type', and 'Study Subject Count'. Each section has a list of checkboxes with study details next to them. At the top right, there's a search bar with filters applied: 'Study Consent IS GRU -- General research use AND Study Subject Count IS More than 10000 AND Study Molecular Data Type IN (SNP/CNV Genotypes (NGS) , Whole Exome (NGS) , Whole Genome (NGS))'. Below the filters, there's a table showing study details, and at the bottom, there are file selection buttons for PubMed, PMC, MeSH, BioProject, BioSample, and SRA.

[https://www.ncbi.nlm.nih.gov/projects/gapsolr/facets.html?COND=%7B%22study_consent_label_abbrev_combo%22:%5B%22GRU%20---%20General%20research%20use%20%22%5D,%22study_subject_count%22:%5B%22More%20than%2010000%22%5D,%22study_sample_use_display_name%22:%5B%22SNP%20Genotypes%20\(NGS\)%22,%22Whole%20Exome%20\(NGS\)%22,%22Whole%20Genome%20\(NGS\)%22%5D%7D](https://www.ncbi.nlm.nih.gov/projects/gapsolr/facets.html?COND=%7B%22study_consent_label_abbrev_combo%22:%5B%22GRU%20---%20General%20research%20use%20%22%5D,%22study_subject_count%22:%5B%22More%20than%2010000%22%5D,%22study_sample_use_display_name%22:%5B%22SNP%20Genotypes%20(NGS)%22,%22Whole%20Exome%20(NGS)%22,%22Whole%20Genome%20(NGS)%22%5D%7D)

✓ Advanced search で疾患, 利用条件, 対象人数などの項目で絞り込みができる

The European Genome-phenome Archive (EGA) is a service for permanent archiving and sharing of all types of personally identifiable genetic and phenotypic data resulting from biomedical research projects.

What is in the EGA?

Studies in the EGA by disease

Click on a column to view category subgroups

Disease Category	Number of studies
Cancer	769
Cardiovascular	122
Infectious	33
Inflammatory	53
Neurological	57
Other	294

If applicable, a study may be included in more than one category

Latest studies

Direct detection of early-stage cancers using circulating tumor – 2017-08-17

Early detection and intervention are likely to be the most effective means for reducing morbidity and mortality of human cancer. However, development of methods for noninvasive detection [Read more →](#)

Study 1 / 4

[Next Study](#)

Published in:

- Various journals
- Genome Biology

BROWSE

- Studies
- Datasets
- DACs
- Beacon

HELP

- FTP & Aspera
- Tools
- EGA Blog
- Contact us
- Twitter

<https://ega-archive.org/>

- ✓ 1,698 studies, 3.5 petabytes のデータ (2017年10月)
- ✓ ストリーミングツールの開発、EU 各国のバイオバンクとの連携



Japanese Genotype-phenotype Archive

Home Studies Submission ▾ FAQ

概要

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

なお、JGA に登録されるデータおよびデータの利用についての審査は独立行政法人科学技術振興機構 (JST)/バイオサイエンスデータベースセンター (NBDC) が実施しています。JGA は科学技術振興機構 National Bioscience Database Center (NBDC) と共同で運営されています。

データの利用

JGA はデータを格納する際にそのデータに適用される利用制限ポリシーを登録しますが、利用者のデータ利用の可否については JST/NBDC が審査します。利用者は NBDC にデータの利用を申請し、JGA は NBDC からの利用承認連絡を受け、利用者にデータへの安全なアクセスを許す形となります。

<http://trace.ddbj.nig.ac.jp/jga/index.html>

Databases ▾ English Contact

Google カスタム検索

データ登録/利用を申請

National Bioscience Database Center

NBDC NBDCヒトデータベース

English

ホーム データの利用 データの提供 ガイドライン NBDCヒトデータ審査委員会 成果発表 お問い合わせ FAQ

NBDCヒトデータベースについて

ヒトに関するデータは、次世代シーケンサーをはじめとした解析技術の発達に伴って膨大な量が産生されつつあり、それらを整理・格納して、生命科学の進展のために有効に活用するためのルールや仕組みが必要です。

国立研究開発法人科学技術振興機構(JST)/バイオサイエンスデータベースセンター(NBDC)では、個人情報の保護に配慮しつつヒトに関するデータの共有や利用を推進するために、ヒトに関する様々なデータを共有するためのプラットフォーム『NBDCヒトデータベース』を設立するとともに、国立遺伝学研究所 DNA Data Bank of Japan & DDBJ と協力して、ヒトに関するデータを公開しています。

本Webサイトを通じて、ヒトに関するデータの利用及びヒトに関するデータの提供を行なうことができます。

なお、本データベースの目的・意義、扱うデータの種類、データ利用者の範囲、責任者についてはこちらをご覧ください。

新着情報

2017/12/26 東京大学大学院 医学系研究科 からの非制限公開データを公開しました (hum0072)

2017/12/26 制限公開データ2件が追加されました (hum0035.v2)

▶ ニュース一覧へ

Search NBDC Human Database Beacon for Alternative Alleles [\[API help\]](#)

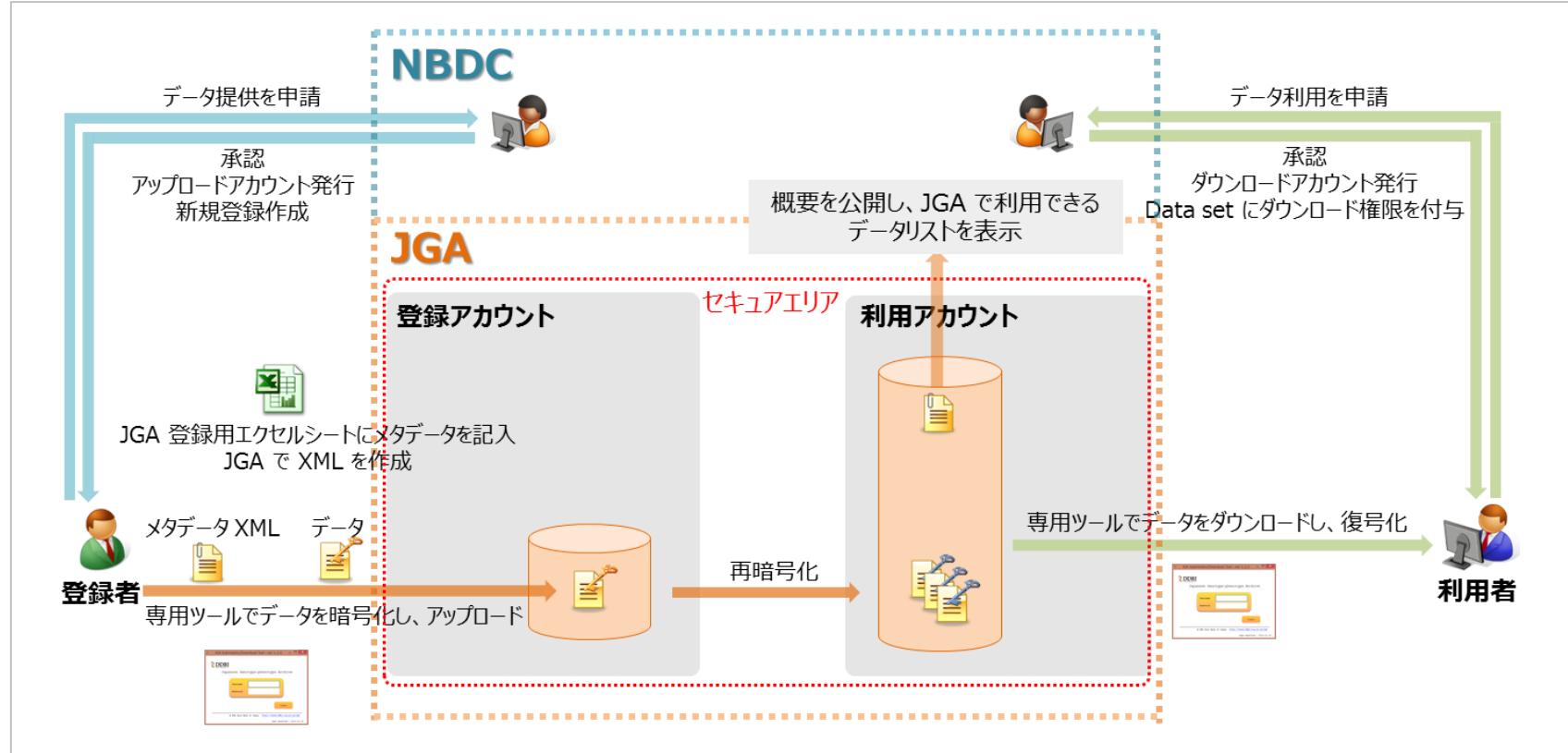
NBDC Human Database Beacon is a member of [GA4GH Beacon Network](#).

GRCh37 e.g. 12:112241766 A Search Example: ALDH2 Variant (GRCh37, '12:112241766 A')

<https://humandbs.biosciencedbc.jp/>

- ✓ 115 studies, 203,345 samples のデータ (2018年1月11日)
- ✓ NBDC ヒトデータ共有ガイドラインに準拠して運営

DDBJ と NBDC の役割分担



<http://trace.ddbj.nig.ac.jp/jga/submission.html>

- ✓ 利用制限が必要な個人由来の遺伝型・表現型情報を受付・保存・提供
- ✓ 原則として匿名化されたメタデータを受付

- ✓ データ提供と利用を NBDC ヒトデータ共有ガイドラインに従って審査

5. 今後の方針性

**NATIONAL CANCER INSTITUTE
GDC Data Portal**

Home Projects Exploration Analysis Repository

Harmonized Cancer Datasets
Genomic Data Commons Data Portal

Get Started by Exploring:

Projects Exploration Analysis Repository

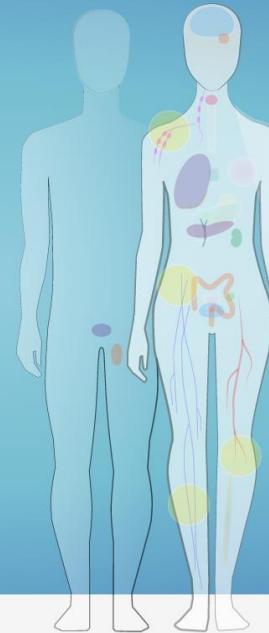
e.g. BRAF, Breast, TCGA-BLCA, TCGA-A5-A0G2

Data Portal Summary Data Release 10 - December 21, 2017

PROJECTS	PRIMARY SITES	CASES
40	60	32,555

FILES	GENES	MUTATIONS
310,859	22,147	3,142,246

Cases by Major Primary Site



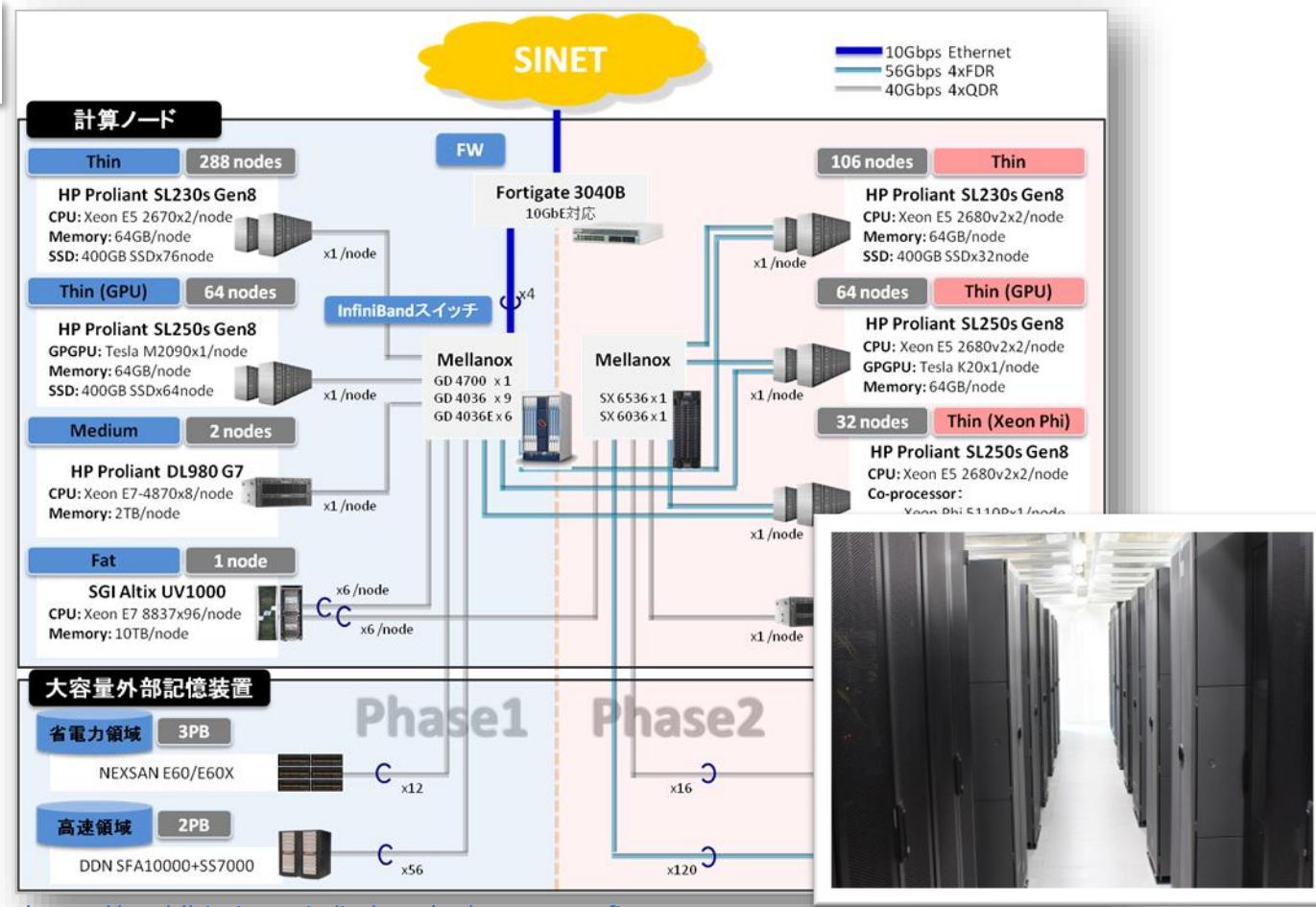

Primary Site	Cases
Adrenal Gland	~100
Bile Duct	~100
Bladder	~1,000
Blood	~1,000
Bone	~100
Bone Marrow	~100
Brain	~1,000
Breast	~3,500
Cervix	~100
Colorectal	~2,800
Esophagus	~500
Eye	~100
Head and Neck	~1,000
Kidney	~2,000
Liver	~1,800
Lung	~3,500
Lymph Nodes	~100
Nervous System	~2,000
Ovary	~1,200
Pancreas	~1,000
Pleura	~100
Prostate	~500
Skin	~1,000
Soft Tissue	~100
Stomach	~1,000
Testis	~100
Thymus	~100
Thyroid	~1,000
Uterus	~1,000

GDC Applications

The GDC Data Portal is a robust data-driven platform that allows cancer researchers and bioinformaticians to search and download cancer data for analysis. The GDC applications include:

 [Data Portal](#)  [Website](#)  [Data Transfer Tool](#)  [API](#)  [Data Submission Portal](#)  [Documentation](#)  [Legacy Archive](#)

- ✓ TCGA データは AWS などのクラウド上で解析できる (認証は NIH eRA アカウント)
データセットのダウンロード不要、解析ツール一式整備済み、強固なセキュリティ



個人ゲノム解析用スパコン

- ✓ 個人ゲノム解析用スパコンの提供準備中
- ✓ スパコンでヒトデータと JGA データの解析ができるようになる