

# バイオインフォマテイクス トレーニング（第3回）

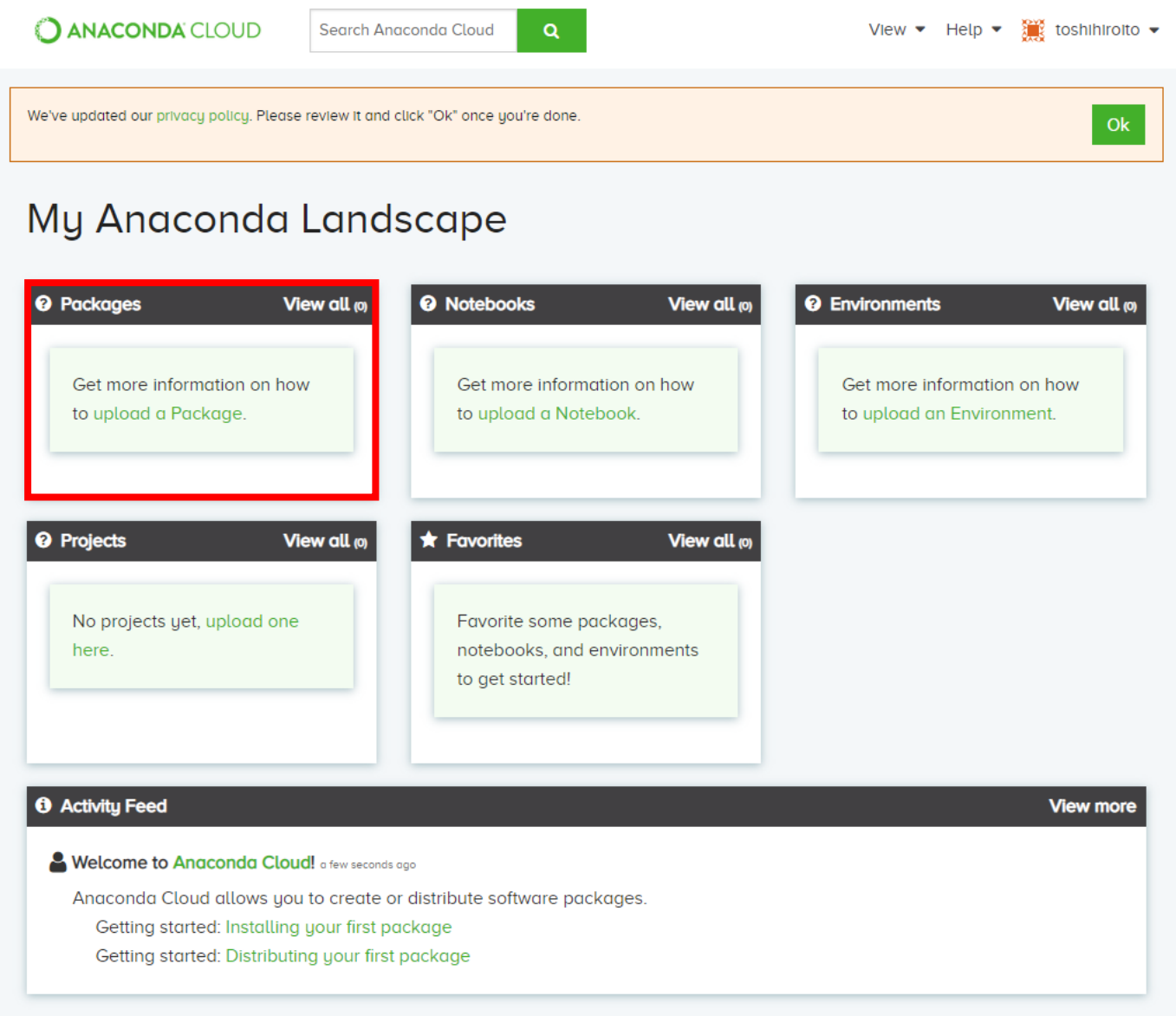
2021年6月24日

国立研究開発法人医薬基盤・健康・栄養研究所

伊藤 寿宏

[toshihiroito@nibiohn.go.jp](mailto:toshihiroito@nibiohn.go.jp)

# Anaconda・Minicondaとは



## ◆ Anaconda・Miniconda

- オープンソースのパッケージマネージャーで、パッケージのインストールから、バージョン管理までを行うことが可能
- Anaconda Cloudでインストール可能なパッケージを確認できる（赤い□の部分）
- Anacondaは、必要以上に多くのパッケージをインストールすることから、インストールに時間がかかる&ディスク容量を圧迫するので、Anacondaの最小構成である**Miniconda**を利用が便利

# Anaconda・Minicondaを用いたパッケージのインストール

2

bioconda / packages / blast 2.9.0



BLAST+ is a new suite of BLAST tools that utilizes the NCBI C++ Toolkit.

Conda

Files

Labels

Badges

License: Public Domain

Home: [http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE\\_TYPE=BlastDocs](http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastDocs)

327854 total downloads

Last upload: 6 months and 2 days ago

## Installers

Info: This package contains files in non-standard labels.

conda install ?

linux-64 v2.9.0

osx-64 v2.9.0

To install this package with conda run one of the following:

```
conda install -c bioconda blast
```

```
conda install -c bioconda/label/cf201901 blast
```

インストールのためにどちらかの  
コマンドを入力

## Description

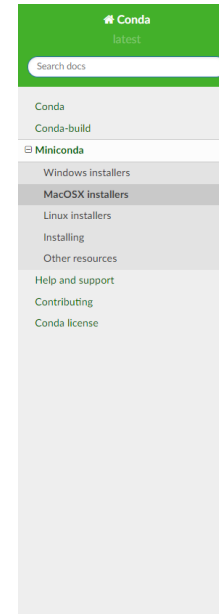
# コマンド操作（Biocondaパッケージをインストールしよう・その1）

3



Bioconda is a channel for the [conda](#) package manager specializing in bioinformatics software. Bioconda consists of:

- a [repository of recipes](#) hosted on GitHub
- a [build system](#) turning these recipes into conda packages
- a [repository of packages](#) containing over 7000 bioinformatics packages ready to use with `conda install`
- over 850 contributors and 570 members who add, modify, update and maintain the recipes



Docs » Miniconda

[Edit on GitHub](#)

## Miniconda

Miniconda is a free minimal installer for conda. It is a small, bootstrap version of Anaconda that includes only conda, Python, the packages they depend on, and a small number of other useful packages, including pip, zlib and a few others. Use the `conda install command` to install 720+ additional conda packages from the Anaconda repository.

See if Miniconda is right for you.

### Windows installers

Windows			
Python version	Name	Size	SHA256 hash
Python 3.7	Miniconda3 Windows 64-bit	51.6 MiB	1781955c063701da05ab4959f6470649b796de73d1578541e5285766405056c
	Miniconda3 Windows 32-bit	52.2 MiB	ca74c06e08731d02972c0f0512629601a8ac3f81ac0331210437201c41f46
Python 2.7	Miniconda2 Windows 64-bit	50.9 MiB	8647c54858f11842c37854e0ff42280c1f5d0802080934676f6a1638664846
	Miniconda2 Windows 32-bit	48.7 MiB	00186228d6461805995f965d50908659329317e30693e327402029197c6f94

### MacOSX installers

MacOSX			
Python version	Name	Size	SHA256 hash
Python 3.7	Miniconda3 MacOSX 64-bit bash	50.3 MiB	d1fc0a474f9971c27208122723843f6c87968d13ff59c01fc817ef62a1576732
	Miniconda3 MacOSX 64-bit pkg	61.3 MiB	f3ede3a58d02f5dc0ba520291a0e0b5c09620840223a20893d540b2758cc00
Python 2.7	Miniconda2 MacOSX 64-bit bash	39.4 MiB	0008f403740e13e102ad0c89e0546f0165478cc770645e720f9c31c8072f39
	Miniconda2 MacOSX 64-bit pkg	47.8 MiB	fcc3002a18f7a292034b2a4a085578a06423f808157f6207794006192f9a0b

## ◆ Minicondaをダウンロードし、システム環境構築しよう（ホームディレクトリ）

wget [weblink](#) #OSおよびPythonのバージョンに合わせて、インストーラーをダウンロードする

【インストーラーを実行】 #minicondaのインストール（色々質問されるが、基本的にEnterを押すか、もしくはYesと入力）

cd miniconda{バージョンの番号}/bin #miniconda2のインストール後に、更新されたbashの設定ファイルを再読み込み

conda -h #condaが正常にインストールされたかどうかチェック。エラーが返ってこなければOK

./conda config --add channels conda-forge

./conda config --add channels defaults

./conda config --add channels r

./conda config --add channels bioconda

} #パッケージのレポジトリ名（チャンネル）をminicondaの設定ファイルに登録する

# condaコマンド

## ◆ conda: ディレクトリ移動のための基本コマンド

- `conda install` 【パッケージ名】 [Anacondaリポジトリからパッケージをインストールする]
- `conda search` 【パッケージ名の一部】 [Anacondaリポジトリに登録されているパッケージを検索する]
- `conda -c` [パッケージをインストールするためのチャンネルを設定する]
- `conda -y` [途中で訊かれる質問をすべてyesで返す]

# コマンド操作（Biocondaパッケージをインストールしよう・その2）

5

- ◆ Biocondaパッケージ”BLAST”をインストールし、さらにNCBIデータベースを構築し、SRR444595.fastqファイルの配列を検索しよう

```
conda install -y htop #システム上で動いているプロセスやCPU・メモリの使用率を閲覧するためのツール
conda install -c bioconda -y blast #問い合わせ配列（Query）に類似した配列（Reference）をデータベース中から検索
するツール。Basic Local Alignment Search Toolの略
conda install -y fastx_toolkit #FASTQ のクオリティコントロールによく利用されるツールの一つ
mkdir -p blastdb/nr && mkdir -p blastdb/nt #NCBIデータベースを構築するためのディレクトリを作成
cd blastdb/nt #ディレクトリ移動
screen #仮想ターミナルの起動
wget ftp://ftp.ncbi.nlm.nih.gov/blast/db/nt.*.tar.gz #NCBI ntデータベースをダウンロード（時間がかかる）
#Ctrl+a+dでディタッチ、プロセス終了後に下記コマンドで解凍
tar -zxvf nt.*.tar.gz

cp ~/SRR444595.fastq ./
fastq_to_fasta -i SRR444595.fastq -o SRR444595.fasta -Q33 #SRR444595.fastqファイルをPhredクオリティスコアが
33（1/2000のベースコールミス）以上のreadだけを残すように、fastaファイルに変換
blastn -db nt -query SRR444595.fasta -out SRR444595.csv #NCBI ntデータベースに対し、SRR444595.fastaファイル
のreadを検索
```

# fasta形式

## ◆ fasta

- NCBIやDDBJなどから塩基配列の情報だけを抽出してシンプルにした形式
  - BLASTを用いて塩基配列、もしくはアミノ酸配列を検索する場合、配列情報をfasta形式で記述する必要がある
- 1行目：>で始まるヘッダ行で、アクセッション番号やシーケンスIDなどが記述されている
  - 2行目：塩基配列、もしくはアミノ酸配列

```
>SRR444595.6 GA-K_0045:1:1:4903:1082/1  
TCTTCACCGGAATGAGAAGGATACTTCAGTTCTGAT
```

# fastq形式

## ◆ fastq

- 次世代シーケンサーで検出されたシーケンスリードの配列を保存した形式
- シーケンスリードのクオリティチェックやデータの加工にはfastq形式のファイルが主に用いられる

- 1行目：@で始まるヘッダ行で、アクセッション番号やシーケンスIDなどが記述されている
- 2行目：塩基配列
- 3行目：+で始まるヘッダ行
- 4行目：配列のクオリティスコア（ASCII 文字で表示）

```
@SRR444595.1 GA-K_0045:1:1:2896:1080/1
GGNGGAGATGTTGTGACTGGTGGTGGTGGCTGAGACA
+
??#9>?BBBBGGGDBDBDDEAFDAFB>BB?E>GG@8
```

Symbol	ASCII Code	Q-Score	Symbol	ASCII Code	Q-Score	Symbol	ASCII Code	Q-Score
!	33	0	/	47	14	=	61	28
"	34	1	0	48	15	>	62	29
#	35	2	1	49	16	?	63	30
\$	36	3	2	50	17	@	64	31
%	37	4	3	51	18	A	65	32
&	38	5	4	52	19	B	66	33
'	39	6	5	53	20	C	67	34
(	40	7	6	54	21	D	68	35
)	41	8	7	55	22	E	69	36
*	42	9	8	56	23	F	70	37
+	43	10	9	57	24	G	71	38
,	44	11	:	58	25	H	72	39
-	45	12	;	59	26	I	73	40
.	46	13	<	60	27			

※塩基配列をシーケシングするときには塩基の解読エラー（ベースコールミス）が生じるので、エラーの生じる確率を $P_{error}$ とすると、クオリティスコア $Q = -10\log_{10}P_{error}$ として算出される（ $Q=10$ : 塩基の解読エラーが生じる確率は10%、 $Q=30$ : 塩基の解読エラーが生じる確率は0.1%）

クオリティスコアとASCII文字は、 $ASCII_{code} = Q + 33 = -10\log_{10}P_{error} + 33$ の関係となる



# NCBI SRA (Sequence Read Archive) へのアクセス

NCBI Resources How To Sign in to NCBI

SRA SRA mRNA-seq mouse white adipose Search

Create alert Advanced Help

**COVID-19 Information**  
[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevention and treatment information \(HHS\)](#) | [Español](#)

Access Public (678)

Source RNA (139)

Library Layout paired (61) single (617)

Platform Illumina (678)

Strategy other (678)

Data in Cloud GS (678) S3 (678)

File Type fastq (140)

[Clear all](#)

[Show additional filters](#)

Summary 20 per page Send to: Filters: [Manage Filters](#)

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

**Search results**  
Items: 1 to 20 of 678

<< First < Prev Page 1 of 34 Next > Last >>

☐ [Illumina HiSeq 4000 sequencing; Single cell RNA-sequencing of lineage-negative fraction of stromal cells isolated from inguinal white adipose tissue in mouse](#)  
1 ILLUMINA (Illumina HiSeq 4000) run: 5.2M spots, 260.5M bases, 122.6Mb downloads  
Accession: ERX3654510

☐ [Illumina HiSeq 4000 sequencing; Single cell RNA-sequencing of lineage-negative fraction of stromal cells isolated from inguinal white adipose tissue in mouse](#)  
1 ILLUMINA (Illumina HiSeq 4000) run: 2.4M spots, 120.6M bases, 71.8Mb downloads  
Accession: ERX3654509

☐ [Illumina HiSeq 4000 sequencing; Single cell RNA-sequencing of lineage-negative fraction of stromal cells isolated from inguinal white adipose tissue in mouse](#)  
1 ILLUMINA (Illumina HiSeq 4000) run: 2.5M spots, 125.6M bases, 61.1Mb downloads  
Accession: ERX3654508

☐ [Illumina HiSeq 4000 sequencing; Single cell RNA-sequencing of lineage-negative fraction of stromal cells isolated from inguinal white adipose tissue in mouse](#)  
1 ILLUMINA (Illumina HiSeq 4000) run: 4.6M spots, 229.4M bases, 106Mb downloads  
Accession: ERX3654507

**Results by taxon**  
Top Organisms [\[Tree\]](#)  
Mus musculus (670)  
Hipposideros armiger (8)

**Search in related databases**

Database	Access		all
	public	controlled	
BioSample	<a href="#">27</a>		<a href="#">27</a>
BioProject	<a href="#">7</a>		<a href="#">7</a>
dbGaP			

Database: Select Find items

**Search details**  
mRNA-seq[All Fields] AND ("Mus"[Organism])

検索したいキーワードを入力

試しにこのプロジェクトをクリック

# データ取得方法

**ERX3654510:** Illumina HiSeq 4000 sequencing; Single cell RNA-sequencing of lineage-negative fraction of stromal cells isolated from inguinal white adipose tissue in mouse

1 ILLUMINA (Illumina HiSeq 4000) run: 5.2M spots, 260.5M bases, 122.6Mb downloads

**Design:** Single cell RNA-sequencing of lineage-negative fraction of stromal cells isolated from inguinal white adipose tissue in mouse

**Submitted by:** UCSF (University of California, San Francisco)

**Study:** Single cell RNA-sequencing of lineage-negative fraction of stromal cells isolated from inguinal white adipose tissue in mouse

[PRJEB35376](#) • [ERP118414](#) • [All experiments](#) • [All runs](#)

[show Abstract](#)

**Sample:** Protocols: Tissue of male C57BL6J mice was harvested, digested, and stromal cells isolated via differential centrifugation. The lineage-negative fraction of stromal cells was purified via affinity column. Following centrifugation and removal of the medium, cells were resuspended at a concentration of 150-500 cells/ $\mu$ L. This cell suspension was mixed with C1 Cell Suspension Reagent (Fluidigm, Cat # 634833) at the recommended ratio of 3:2 immediately before loading 5  $\mu$ L of this final mix on the C1 IFC. Cells were lysed via Module 2 mRNA-seq kit per Fluidigm protocol specifications (PN 100-7168). Resultant mRNA was stabilized and reverse-transcribed using the SMARTer kit (Clontech), also per Fluidigm protocol specifications. Approximate run time was ~7.75 hours (6.5 hours for lysis, reverse transcription, and amplification; and 1.25 hours for harvest). 3 $\mu$ L amplicons were diluted with 10 $\mu$ L C1 DNA Dilution Reagent (Fluidigm) to generate 13  $\mu$ L mix, and transferred to a new plate. Sample concentrations were normalize

[SAMEA6225556](#) • [ERS4025322](#) • [All experiments](#) • [All runs](#)

**Organism:** [Mus musculus](#)

**Library:**

**Name:** ING\_(no-Tx)\_single-219\_s

**Instrument:** Illumina HiSeq 4000

**Strategy:** RNA-Seq

**Source:** TRANSCRIPTOMIC SINGLE CELL

**Selection:** Oligo-dT

**Layout:** SINGLE

**Construction protocol:** Tissue of male C57BL6J mice was harvested, digested, and stromal cells isolated via differential centrifugation. The lineage-negative fraction of stromal cells was purified via affinity column. Following centrifugation and removal of the medium, cells were resuspended at a concentration of 150–500 cells/ $\mu$ L. This cell suspension was mixed with C1 Cell Suspension Reagent (Fluidigm, Cat # 634833) at the recommended ratio of 3:2 immediately before loading 5  $\mu$ L of this final mix on the C1 IFC. Cells were lysed via Module 2 mRNA-seq kit per Fluidigm protocol specifications (PN 100-7168). Resultant mRNA was stabilized and reverse-transcribed using the SMARTer kit (Clontech), also per Fluidigm protocol specifications. Approximate run time was ~7.75 hours (6.5 hours for lysis, reverse transcription, and amplification; and 1.25 hours for harvest). 3 $\mu$ L amplicons were diluted with 10 $\mu$ L C1 DNA Dilution Reagent (Fluidigm) to generate 13  $\mu$ L mix, and transferred to a new plate. Sample concentrations were normalized manually, then prepared for successive tagmentation, amplification, and pooling per the Illumina Nextera XT DNA Library Preparation protocol guidelines.



**Runs:** 1 run, 5.2M spots, 260.5M bases, [122.6Mb](#)

Run	# of Spots	# of Bases	Size	Published
<a href="#">ERR3662870</a>	5,209,918	260.5M	122.6Mb	2020-04-29



# データ取得方法

Illumina HiSeq 4000 sequencing; Single cell RNA-sequencing of lineage-negative fraction of stromal cells isolated from inguinal white adipose tissue in mouse (ERR3662870)

Metadata Analysis Reads Data access

## SRA archive data

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

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

Type	Size	Location	Name	Free Egress	Access Type
run	125,544 Kb	NCBI	<a href="https://sra-download.ncbi.nlm.nih.gov/traces/era9/ERR/ERR3662/ERR3662870">https://sra-download.ncbi.nlm.nih.gov/traces/era9/ERR/ERR3662/ERR3662870</a>	worldwide	anonymous
		AWS	<a href="https://sra-pub-run-odp.s3.amazonaws.com/sra/ERR3662870/ERR3662870">https://sra-pub-run-odp.s3.amazonaws.com/sra/ERR3662870/ERR3662870</a>	worldwide	anonymous
		GCP	gs://sra-pub-run-6/ERR3662870/ERR3662870.1	gs.US	gcp identity

ここからダウンロード

◆ SRA形式なので、fastqに変換して使用

# BLAST (Basic Local Alignment Search Tool)

- ◆ BLASTは、入力配列 (**query**) と相同性のある配列 (**subject**) をデータベースから検索するツール (**Web版・ローカル版**がある)
  - 入力配列、もしくは検索対象となるデータベースの種類 (核酸・アミノ酸) の組み合わせによりblastのプログラム名が異なる
- 今回は”Nucleotide BLAST (核酸→核酸)”を使用する

**Basic Local Alignment Search Tool**

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)

**Search Betacoronavirus Database**

We have created a new BLAST database focused on the SARS-CoV-2 (Severe acute respiratory syndrome coronavirus 2) Sequences. For further detail please visit [NCBI GenBank](#).

Mon, 03 Feb 2020 10:00:00 EST [More BLAST news...](#)

**Web BLAST**

**Nucleotide BLAST**  
nucleotide ▶ nucleotide

**blastx**  
translated nucleotide ▶ protein

**tblastn**  
protein ▶ translated nucleotide

**Protein BLAST**  
protein ▶ protein

核酸→核酸      核酸→アミノ酸      アミノ酸→核酸      アミノ酸→アミノ酸

# BLAST (Web版blastn・解析例)

## Standard Nucleotide BLAST

blastn [blastp](#) [blastx](#) [tblastn](#) [tblastx](#)

BLASTN programs search nucleotide databases using a nucleotide query. [more...](#)

[Reset page](#) [Bookmark](#)

### Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#)

>SRR444595.6 GA-K\_0045:1:1:4903:1082/1  
TCTTCACCGGAATGAGAGGATCTTCAGTTCTGAT

[Clear](#) Query subrange [?](#)

From

To

Or, upload file [?](#)

[ファイルを選択](#) VZV\_UT.fasta [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

### Choose Search Set

Database ☒ Standard databases (nr etc.): ☐ rRNA/ITS databases ☐ Genomic + transcript databases ☐ Betacoronavirus

Nucleotide collection (nr/nt) [?](#)

Organism [?](#)

Optional

Enter organism name or id—completions will be suggested ☐ exclude [+](#)

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Exclude [?](#)

Optional ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Limit to [?](#)

Optional ☐ Sequences from type material

Entrez Query [?](#)

Optional

[YouTube](#) [Create custom database](#)

Enter an Entrez query to limit search [?](#)

### Program Selection

Optimize for

☒ Highly similar sequences (megablast)

☐ More dissimilar sequences (discontiguous megablast)

☐ Somewhat similar sequences (blastn)

Choose a BLAST algorithm [?](#)

[BLAST](#) [Search database](#) [Show results in](#)

検索開始

nt) using Megablast (Optimize for highly similar sequences)

BLAST results will be displayed in a new format by default. You can always switch back to the Traditional Results page. [New](#)

検索したい配列を直接入力、もしくは  
fasta形式でアップロード

データベース"Database"、検索  
したいもしくは検索対象外とした  
い生物種"Organism"を選択

検索エンジンの選択:  
正確性 (megablast > discontiguous megablast > blastn)  
網羅性 (blastn > discontiguous megablast > megablast)



# BLAST (Web版blastn・解析例その2)

Job Title	SRR444595.6 GA-K_0045:1:1:4903:1082/1
RID	<a href="#">DTC8JYNU014</a> Search expires on 06-08 17:19 pm <a href="#">Download All</a> ▼
Program	BLASTN <a href="#">?</a> <a href="#">Citation</a> ▼
Database	nt <a href="#">See details</a> ▼
Query ID	lcl Query_48423
Description	SRR444595.6 GA-K_0045:1:1:4903:1082/1
Molecule type	dna
Query Length	36
Other reports	<a href="#">Distance tree of results</a> <a href="#">MSA viewer</a> <a href="#">?</a>

## Filter Results

**Organism** *only top 20 will appear* ☐ exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity  to  E value  to  Query Coverage  to

[Filter](#) [Reset](#)

## ● Query Cover

Query配列の使用率。  
この値が小さい場合、query配列の一部分だけが類似していることを示す。

## ● E value

配列の一致が偶然現れる期待値。  
統計検定におけるp値と同じような意味で、値が小さいほど偶然では起こりえない一致を示す。

## ● Per. Ident

Query配列とデータベース配列の一致率。

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignments

Download

Manage Columns

Show

100

☒ select all
 20 sequences selected

[GenBank](#)
[Graphics](#)
[Distance tree of results](#)

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	<a href="#">Arabidopsis thaliana genome assembly, chromosome: 1</a>	67.6	67.6	100%	8e-09	100.00%	<a href="#">LR782542.1</a>
<input checked="" type="checkbox"/>	<a href="#">Arabidopsis thaliana genome assembly, chromosome: 1</a>	67.6	67.6	100%	8e-09	100.00%	<a href="#">LR699745.2</a>
<input checked="" type="checkbox"/>	<a href="#">Arabidopsis thaliana genome assembly, chromosome: 1</a>	67.6	67.6	100%	8e-09	100.00%	<a href="#">LR699770.1</a>
<input checked="" type="checkbox"/>	<a href="#">Arabidopsis thaliana genome assembly, chromosome: 1</a>	67.6	67.6	100%	8e-09	100.00%	<a href="#">LR699765.1</a>
<input checked="" type="checkbox"/>	<a href="#">Arabidopsis thaliana genome assembly, chromosome: 1</a>	67.6	67.6	100%	8e-09	100.00%	<a href="#">LR699760.1</a>
<input checked="" type="checkbox"/>	<a href="#">Arabidopsis thaliana genome assembly, chromosome: 1</a>	67.6	67.6	100%	8e-09	100.00%	<a href="#">LR699755.1</a>
<input checked="" type="checkbox"/>	<a href="#">Arabidopsis thaliana genome assembly, chromosome: 1</a>	67.6	67.6	100%	8e-09	100.00%	<a href="#">LR699750.1</a>
<input checked="" type="checkbox"/>	<a href="#">Arabidopsis thaliana genome assembly, chromosome: 1</a>	67.6	67.6	100%	8e-09	100.00%	<a href="#">LR215052.1</a>
<input checked="" type="checkbox"/>	<a href="#">Arabidopsis thaliana Ribosomal protein L16p/L10e family protein (SAC52), mRNA</a>	67.6	67.6	100%	8e-09	100.00%	<a href="#">NM_101298.3</a>
<input checked="" type="checkbox"/>	<a href="#">Arabidopsis thaliana Ribosomal protein L16p/L10e family protein (SAC52), mRNA</a>	67.6	67.6	100%	8e-09	100.00%	<a href="#">NM_001332111.1</a>
<input checked="" type="checkbox"/>	<a href="#">Arabidopsis thaliana chromosome 1 sequence</a>	67.6	67.6	100%	8e-09	100.00%	<a href="#">CP002684.1</a>
<input checked="" type="checkbox"/>	<a href="#">Arabidopsis thaliana clone RAFL07-10-D02 (R10643) putative tumor suppressor (At1g14320) mRNA, complete cds</a>	67.6	67.6	100%	8e-09	100.00%	<a href="#">BT000679.1</a>
<input checked="" type="checkbox"/>	<a href="#">Arabidopsis thaliana putative tumor suppressor protein (At1g14320) mRNA, complete cds</a>	67.6	67.6	100%	8e-09	100.00%	<a href="#">AY045866.1</a>
<input checked="" type="checkbox"/>	<a href="#">Arabidopsis thaliana At1g14320/F14L17_28 mRNA, complete cds</a>	67.6	67.6	100%	8e-09	100.00%	<a href="#">AF428470.1</a>

# BLAST (Web版blastn・解析例その3)

[Download](#) [GenBank](#) [Graphics](#)

[Next](#) [Previous](#) [Descriptions](#)

**Arabidopsis thaliana genome assembly, chromosome: 1**

Sequence ID: [LR782542.1](#) Length: 30272351 Number of Matches: 1

Range 1: 4893844 to 4893879 [GenBank](#) [Graphics](#)

[Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
67.6 bits(36)	8e-09	36/36(100%)	0/36(0%)	Plus/Minus
Query 1	TCTTCACCGGAATGAGAAGGATACTTCAGTTCTGAT 36			
Sbjct 4893879	TCTTCACCGGAATGAGAAGGATACTTCAGTTCTGAT 4893844			

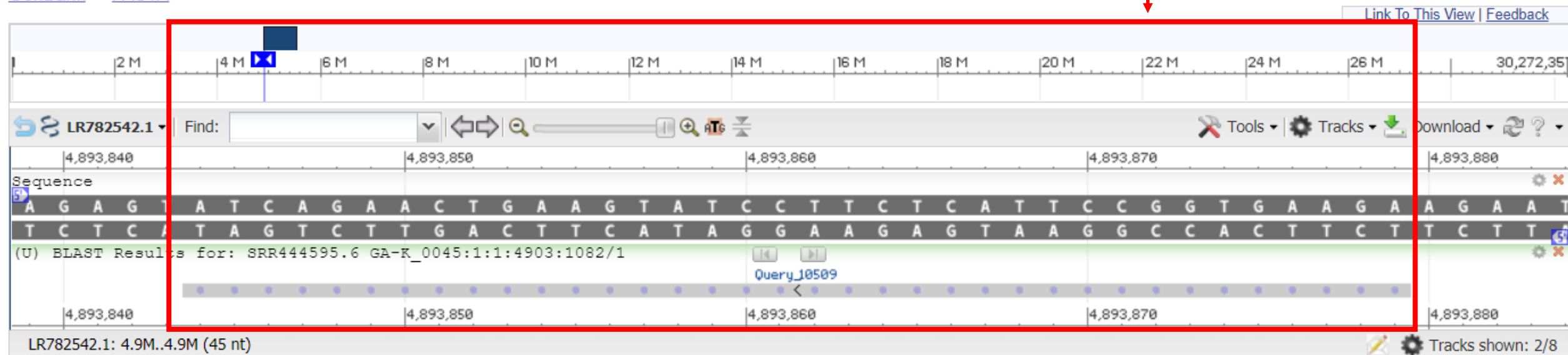
構造情報：query配列/database配列  
Plus (塩基配列)、Minus (相補配列)

**Arabidopsis thaliana genome assembly, chromosome: 1**

GenBank: [LR782542.1](#)

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

位置情報



補足資料



Table 1 ASCII Characters Encoding Q-scores 0-40

Symbol	ASCII Code	Q-Score	Symbol	ASCII Code	Q-Score	Symbol	ASCII Code	Q-Score
!	33	0	/	47	14	=	61	28
"	34	1	0	48	15	>	62	29
#	35	2	1	49	16	?	63	30
\$	36	3	2	50	17	@	64	31
%	37	4	3	51	18	A	65	32
&	38	5	4	52	19	B	66	33
'	39	6	5	53	20	C	67	34
(	40	7	6	54	21	D	68	35
)	41	8	7	55	22	E	69	36
*	42	9	8	56	23	F	70	37
+	43	10	9	57	24	G	71	38
,	44	11	:	58	25	H	72	39
-	45	12	;	59	26	I	73	40
.	46	13	<	60	27			

**ASCII（アスキー、英: American Standard Code for Information Interchange）**は、現代英語や西ヨーロッパ言語で使われるラテン文字を中心とした文字コード。これはコンピュータその他の通信機器において最もよく使われているものである（Wikipediaより）。