



# NIG Bioinformatics Training Program

Eli Kaminuma (National Institute of Genetics)

2015-11-25

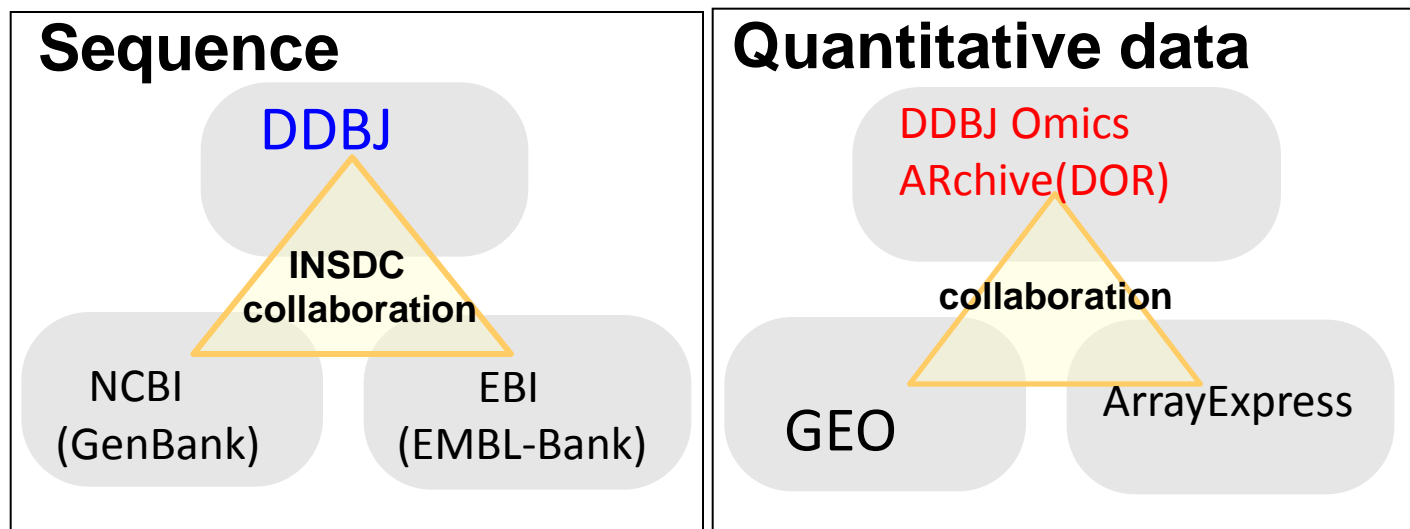
## 1. SRA database

- **Biosample**
- **SRA datamodel**

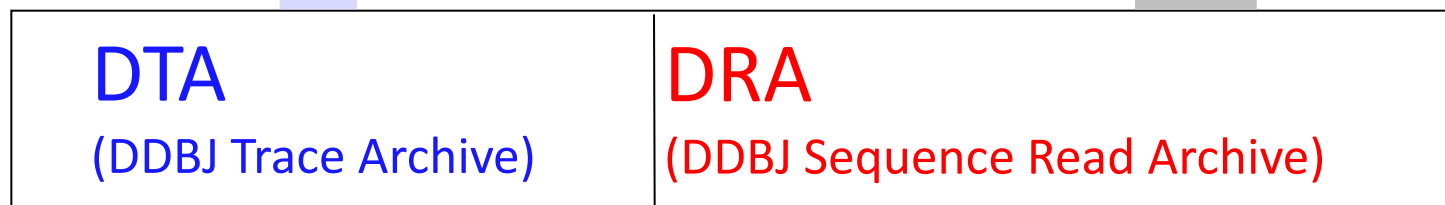


# 1-1) DDBJ databases for next-generation sequencing data with the international collaboration

Archive DB for analyzed data

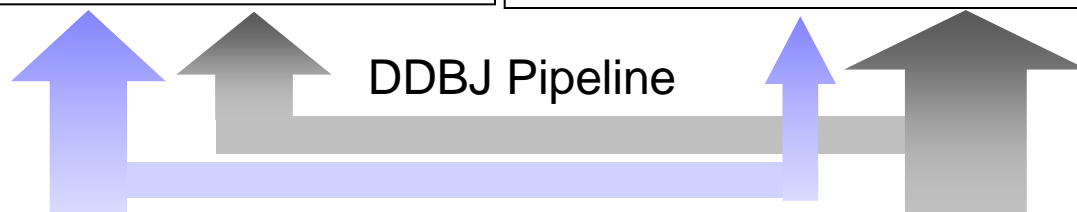


Archive DB for raw data



Capillary Electrophoresis Sequencer

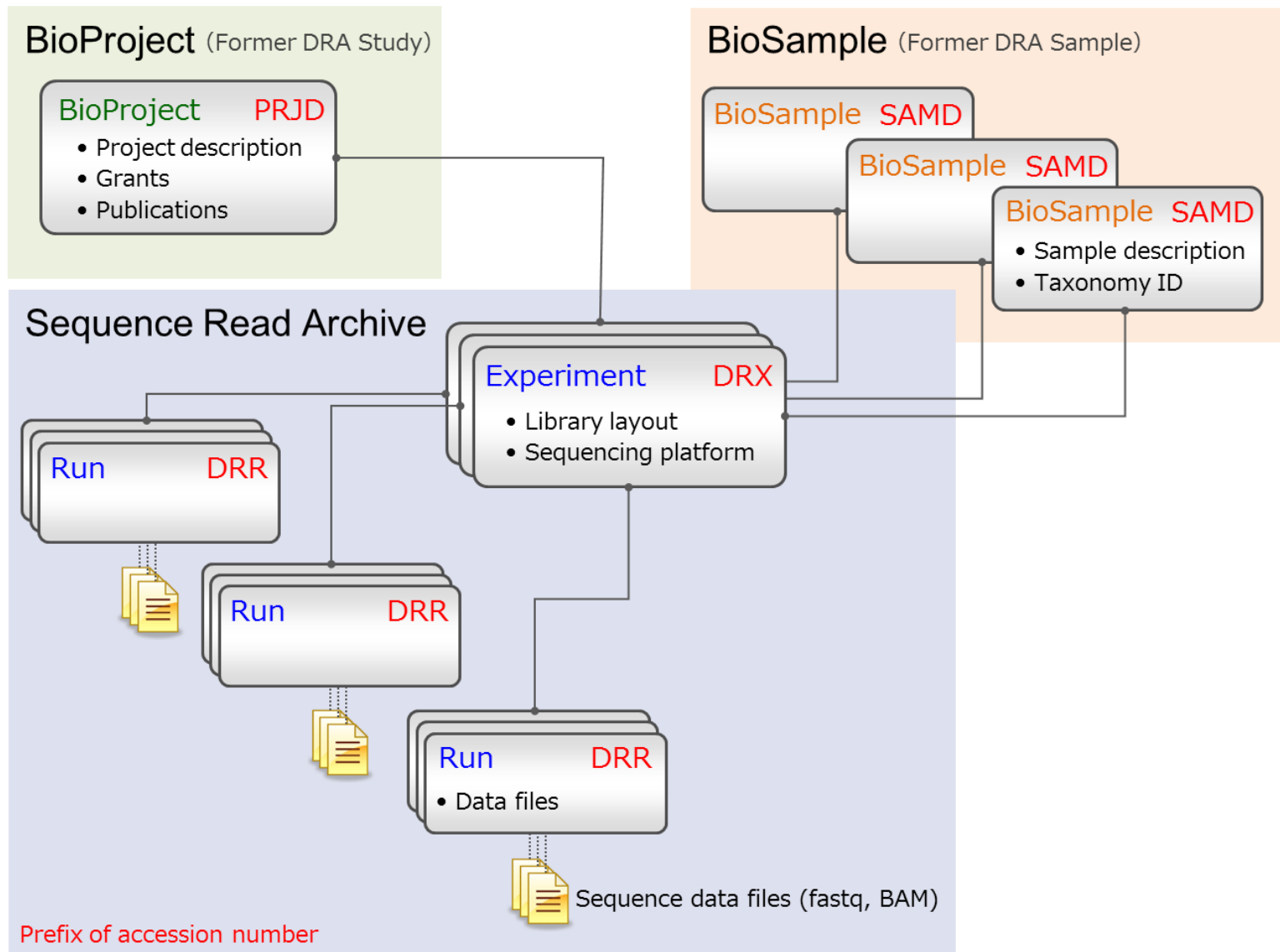
Next-generation Sequencer (NGS)





## 1-2) SRA database : Data model

[http://trace.ddbj.nig.ac.jp/dra/submission\\_e.html](http://trace.ddbj.nig.ac.jp/dra/submission_e.html)





# 1-3) DDBJ-SRA search (1)

## DRASearch

☒ Send Feedback [Se](#)

Accession :

Organism :  StudyType :

CenterName :  Platform :

Keyword :  (1) Input keyword

Show  records Sort by  (2) Click search button

Search Results ( 71 records )

<< 

### Filtered by

(3) Select 'Cocos nucifera'

document type:sample(25) experiment(17) study(16) run(7) submission(6)  
organism:Nelumbo nucifera(25)  Wolfiporia cocos MD-104 SS10(7)  
Wolfiporia cocos CGMCC 5.0078(4) marine metagenome(2) Afipia felis(1)

| # | META_FILE | ACCESSION | STUDY | STUDY_TITLE | STUDY_TYPE | ORGANISM | BASES |
|---|-----------|-----------|-------|-------------|------------|----------|-------|
|---|-----------|-----------|-------|-------------|------------|----------|-------|

## DRASearch

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

Accession :

Organism :  StudyType :

CenterName :  Platform :

Keyword :

Show  records Sort by

Search Results ( 20 records )

<< 

### Filtered by

document type:sample(8) run(5) experiment(4) submission(3)  
organism:Cocos nucifera

| # | META_FILE  | ACCESSION                              | STUDY     | STUDY_TITLE  | STUDY_TYPE             | ORGANISM       | BASES | SUBMITTED | CENTER_NAME |
|---|--|--|-----------|--|------------------------|----------------|-------|-----------|-------------|
| 1 | SRA161945.submission.xml<br><?xml version="1.0" encoding="UTF-8"?><br><SUBMISSION alias="Cocos nucifera transcriptome" lab                           | <input type="text" value="SRA161945"/> | SRP041201 | Cocos nucifera Transcriptome or Gene expression            | Transcriptome Analysis | Cocos nucifera | 39.2G |           | BioProject  |
| 2 | SRA128266.submission.xml<br><?xml version="1.0" encoding="UTF-8"?><br><SUBMISSION alias="Cocos nucifera root (wild) disease healthy                  | SRA128266                              | SRP035487 | Cocos nucifera Transcriptome or Gene expression            | Other                  | Cocos nucifera | 22.4G |           | BioProject  |
| 3 | SRA140250.sample.xml<br>.org/2001/XMLSchema-instance"><br><SAMPLE center_name="CPCRI" alias="COCOS NUCIFERA EMBRYOGENIC CALLUS" accession="SR5558711 | SR5558711                              | SRP035511 | Cocos nucifera strain:WCT Transcriptome or Gene expression | Other                  | Cocos nucifera | 31.8G |           | BioProject  |
| 4 | SRA105100.submission.xml<br><?xml version="1.0" encoding="UTF-8"?><br><SUBMISSION alias="CNU" submission_comment="Cocos nucifera                     | SRA105100                              | SRP035511 | Cocos nucifera strain:WCT Transcriptome or Gene expression | Other                  | Cocos nucifera | 31.8G |           | BioProject  |

(5) Click to the next

(4) 20 dataset retrieved



# 1-4) DDBJ-SRA search(2)

DRASearch

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

SRA161945

[FTP](#)

(6)Click to the next

## Submission Detail

|                 |                              |
|-----------------|------------------------------|
| Alias           | Cocos nucifera transcriptome |
| Submission ID   |                              |
| Submission Date |                              |
| Center Name     | Academia Sinica              |
| Lab Name        | Matzke group                 |

## Navigation

|            |  |
|------------|--|
| Study      | <a href="#">SRP041201</a>  |
| Experiment | <a href="#">SRX534380</a> <a href="#">FASTQ</a> <a href="#">SRA</a>  |
|            | <a href="#">SRX534428</a> <a href="#">FASTQ</a> <a href="#">SRA</a>  |
| Sample     | <a href="#">SRS603105</a>  |
|            | <a href="#">SRS603153</a>  |
| Run        | <a href="#">SRR1273070</a> <a href="#">FASTQ</a> <a href="#">SRA</a> |
|            | <a href="#">SRR1273180</a> <a href="#">FASTQ</a> <a href="#">SRA</a> |

Website policy | © DNA Data Bank of Japan

## /ddbj\_database/dra/fastq/SRA161/SRA161945 のインデックス

| 名前                                       | サイズ    |
|--|--------|
| <a href="#">[親ディレクトリ]</a>                |        |
| <a href="#">SRA161945.experiment.xml</a> | 6.0 kB |
| <a href="#">SRA161945.run.xml</a>        | 1.3 kB |
| <a href="#">SRA161945.sample.xml</a>     | 2.7 kB |
| <a href="#">SRA161945.submission.xml</a> | 220 B  |

(7)Click to the next

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```

<SAMPLE_SET xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <SAMPLE center_name="Academia Sinica" alias="Taro dwarf coconut embryo" accession="SRS603105">
    <IDENTIFIERS>
      <PRIMARY_ID>SRS603105</PRIMARY_ID>
      <EXTERNAL_ID namespace="BioSample">SAMN02728104</EXTERNAL_ID>
      <SUBMITTER_ID namespace="Academia Sinica">Taro dwarf coconut embryo</SUBMITTER_ID>
    </IDENTIFIERS>
    <TITLE>Cocos nucifera embryo</TITLE>
    <SAMPLE_NAME>
      <TAXON_ID>13894</TAXON_ID>
      <SCIENTIFIC_NAME>Cocos nucifera</SCIENTIFIC_NAME>
    </SAMPLE_NAME>
    <SAMPLE_ATTRIBUTES>
      <SAMPLE_ATTRIBUTE>
        <TAG>cultivar</TAG>
        <VALUE>dwarf cultivar</VALUE>
      </SAMPLE_ATTRIBUTE>
      <SAMPLE_ATTRIBUTE>
        <TAG>biomaterial_provider</TAG>
        <VALUE>Matzke group, Institute of Plant and Microbial Biology, Academia Sinica, Taipei 11529, Taiwan</VALUE>
      </SAMPLE_ATTRIBUTE>
    </SAMPLE_ATTRIBUTES>
  </SAMPLE>
</SAMPLE_SET>

```

(8)Sample information



## ASSIGNMENT[1]

### 1-4) Extracting biosample conditions from SRA metadata

ORGANISM=**Cocos nucifera**

<TAXON\_ID>**13894**</TAXON\_ID>

| ACCESSION | STUDY     | EXPERIMENT | RUN        | SAMPLE    | <TAG>cultivar</TAG> | <TAG>tissue</TAG> | Library Strategy |
|-----------|-----------|------------|------------|-----------|---------------------|-------------------|------------------|
| SRA161945 | SRP041201 | SRX534380  | SRR1273070 | SRS603105 | dwarf cultivar      | embryo            | RNA-Seq          |
|           | SRP041201 | SRX534428  | SRR1273180 | SRS603153 | dwarf cultivar      | young leaf        | RNA-Seq          |
| SRA060366 |           |            |            |           |                     |                   |                  |
| SRA122562 |           |            |            |           |                     |                   |                  |
| SRA128266 |           |            |            |           |                     |                   |                  |
| SRA128321 |           |            |            |           |                     |                   |                  |
| SRA157871 |           |            |            |           |                     |                   |                  |

## ASSIGNMENT[1]

Fill required entries based on metadata of 6 SRA datasets for *cocos nucifera*

SRA161945 - SRA157871



## ASSIGNMENT[2]

### 1-5) Extracting biosample conditions from SRA metadata

#### ASSIGNMENT[2]

Fill required entries based on metadata of 6 SRA datasets for *Cocos nucifera* by programming

ORGANISM=**Cocos nucifera**

<TAXON\_ID>**13894**</TAXON\_ID>

SRA ID: **SRA161945, SRA060366, SRA122562, SRA128266, SRA128321, SRA157871**

Retrieval SRA metadata in NIG supercomputer

```
[kaminuma@t262 ~]$ ssh t347 (SRA data server)
[kaminuma@t347 ~]$
[kaminuma@t347 ~]$ mkdir sraxml
:
[kaminuma@t347 SRA161945]$ cd
    /usr/local/ftp/public/ddbj_database/dra/fastq/SRA161/SRA161945
[kaminuma@t347 SRA161945]$ ls
./ SRA161945.experiment.xml SRA161945.sample.xml
../ SRA161945.run.xml      SRA161945.submission.xml
[kaminuma@t347 SRA161945]$ cp * ~/sraxml/
```



# Reference 1: SRA database in NIG supercomputer

- ssh \*\*
- qlogin
- ssh t347
- cd /usr/local/ftp/public/ddbj\_database/dra/

t347 : SRAデータベースにアクセスできるノード

```
[kaminuma@t347 dra]$ ls -lt
合計 31
drwxrwxr-x 612 tracesys tracesys 612 2月 21 18:01 2015 fastq/
drwxrwxr-x 810 tracesys tracesys 810 2月 13 22:00 2015 DRA000/
drwxr-xr-x+ 20 dbm dbm 21 1月 9 13:45 2015 ../
drwxr-xr-x+ 3 tracesys tracesys 3 9月 10 11:31 2014 sra/
drwxr-xr-x+ 3 dbm dbm 3 5月 21 14:47 2013 sralite/
drwxr-xr-x+ 9 dbm dbm 9 2月 13 18:43 2012 ./
drwxr-xr-x+ 3 tracesys tracesys 3 12月 20 09:05 2011 meta/
drwxr-xr-x+ 3 tracesys tracesys 8 11月 18 16:34 2011 DRA000001/
drwxr-xr-x+ 3 tracesys tracesys 8 11月 18 16:34 2011 DRA000010/
```

- cd fastq 下が、SRAデータ置き場

```
ERA052/ ERA143/ ERA214/ ERA289/ ERA360/ SRA021/ SRA091/ SRA163/ SRA231/
ERA053/ ERA144/ ERA215/ ERA290/ ERA361/ SRA022/ SRA092/ SRA164/ SRA232/
ERA061/ ERA145/ ERA216/ ERA291/ ERA362/ SRA023/ SRA093/ SRA165/ SRA233/
ERA062/ ERA146/ ERA217/ ERA292/ ERA363/ SRA024/ SRA094/ SRA166/ SRA234/
ERA063/ ERA147/ ERA218/ ERA293/ ERA364/ SRA025/ SRA095/ SRA167/ SRA235/
ERA064/ ERA148/ ERA219/ ERA294/ ERA365/ SRA026/ SRA096/ SRA168/ SRA236/
ERA065/ ERA149/ ERA221/ ERA295/ ERA366/ SRA027/ SRA097/ SRA169/ SRA237/
ERA066/ ERA150/ ERA222/ ERA296/ ERA367/ SRA028/ SRA098/ SRA170/ SRA238/
ERA067/ ERA151/ ERA223/ ERA297/ ERA368/ SRA029/ SRA099/ SRA171/ SRA239/
ERA068/ ERA152/ ERA224/ ERA298/ ERA369/ SRA030/ SRA100/ SRA172/ SRA240/
ERA069/ ERA153/ ERA225/ ERA299/ ERA370/ SRA031/ SRA101/ SRA173/ SRA241/
ERA070/ ERA154/ ERA226/ ERA300/ ERA371/ SRA032/ SRA102/ SRA174/ SRA242/
ERA071/ ERA155/ ERA227/ ERA301/ ERA372/ SRA033/ SRA103/ SRA175/ SRA243/
```

- cd meta/list 下のfastqlistが、SRA全リスト

```
[kaminuma@t347 list]$ ls -lt
合計 238857
-rw-rw-r--+ 1 tracesys tracesys 112180908 2月 27 02:30 2015 sralist
-rw-rw-r--+ 1 tracesys tracesys 36697 2月 27 02:00 2015 centerlist
-rw-rw-r--+ 1 tracesys tracesys 125159510 2月 27 01:31 2015 fastqlist
-rw-rw-r--+ 1 tracesys tracesys 6744731 2月 26 22:07 2015 livelist
drwxrwxr-x 2 tracesys tracesys 7 9月 9 13:44 2014 ./
drwxr-xr-x+ 3 tracesys tracesys 3 12月 20 09:05 2011 ../
-rw-rw-r--+ 1 tracesys tracesys 74335 12月 5 13:18 2011 datalist
```





## Reference 2: SRA example -- DRA0000001

/usr/local/ftp/public/ddbj\_database/dra/fastq/DRA000/DRA0000001

```
合計 24
drwxrwxr-x 810 tracesys tracesys 810 2月 13 22:00 2015 ./
-rw-r--r--+ 1 tracesys tracesys 2947 1月 30 10:56 2015 DRA0000001.experiment.xml
-rw-r--r--+ 1 tracesys tracesys 1002 1月 30 10:56 2015 DRA0000001.sample.xml
-rw-r--r--+ 1 tracesys tracesys 2051 1月 30 10:56 2015 DRA0000001.study.xml
drwxrwxr-x 2 tracesys tracesys 5 6月 6 14:43 2014 DRX0000001/
-rw-r--r--+ 1 tracesys tracesys 377 5月 12 22:06 2014 DRA0000001.submission.xml
-rw-r--r--+ 1 tracesys tracesys 323 10月 31 11:37 2012 DRA0000001.run.xml
```

### DRX0000001 contents

```
合計 806623
drwxrwxr-x 2 tracesys tracesys 5 6月 6 14:43 2014 ./
drwxr-xr-x+ 3 tracesys tracesys 8 11月 18 16:34 2011 ../
-rw-r--r--+ 1 tracesys tracesys 1387158 6月 6 14:37 2014 DRR0000001.fastq.bz2
-rw-r--r--+ 1 tracesys tracesys 399945072 6月 6 14:37 2014 DRR0000001_1.fastq.bz2
-rw-r--r--+ 1 tracesys tracesys 423875917 6月 6 14:37 2014 DRR0000001_2.fastq.bz2
```

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<SAMPLE_SET>
  <SAMPLE accession="DRS0000001" center_name="KEIO" alias="DRS0000001">
    <IDENTIFIERS>
      <PRIMARY_ID label="BioSample ID">SAMD00016353</PRIMARY_ID>
    </IDENTIFIERS>
    <TITLE>DRS0000001</TITLE>
    <SAMPLE_NAME>
      <TAXON_ID>645657</TAXON_ID>
      <SCIENTIFIC_NAME>Bacillus subtilis subsp. natto BEST195</SCIENTIFIC_NAME>
    </SAMPLE_NAME>
    <DESCRIPTION>Genomic DNA from Bacillus subtilis subsp. natto BEST195 plasmid.</DESCRIPTION>
    <SAMPLE_ATTRIBUTES>
      <SAMPLE_ATTRIBUTE>
        <TAG>strain</TAG>
        <VALUE>BEST195</VALUE>
      </SAMPLE_ATTRIBUTE>
      <SAMPLE_ATTRIBUTE>
        <TAG>common name</TAG>
        <VALUE>Bacillus subtilis subsp. natto</VALUE>
        <UNITS></UNITS>
      </SAMPLE_ATTRIBUTE>
    </SAMPLE_ATTRIBUTES>
  </SAMPLE>
</SAMPLE_SET>
```

taxon\_id and strain name  
in \*.sample.xml



# NIG Bioinformatics Training Program

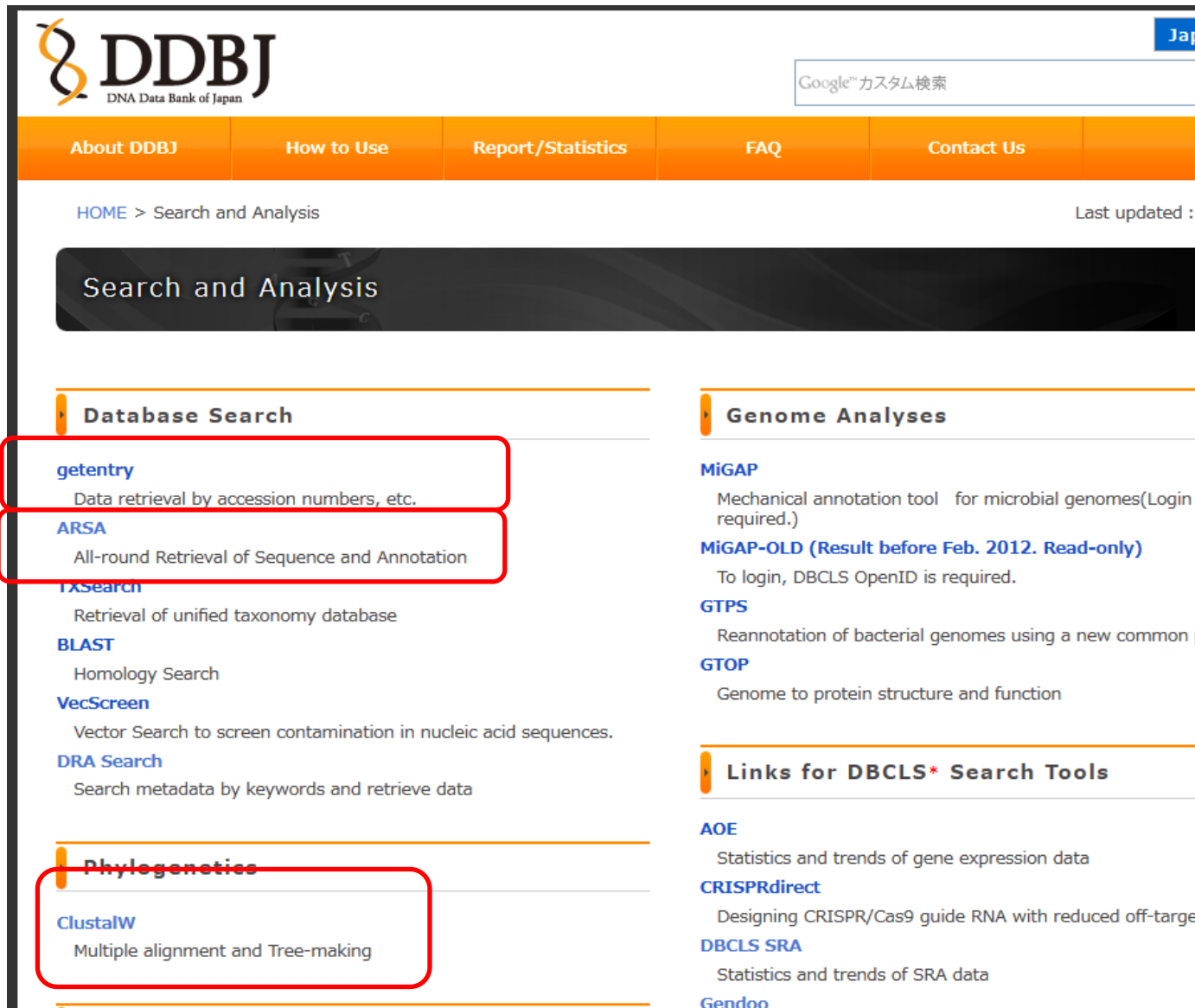
Eli Kaminuma (National Institute of Genetics)

2015-11-25

## 2. DDBJ web tools

- getentry
- ARSA
- ClustalW

<http://www.ddbj.nig.ac.jp/searches-e.html>



The screenshot shows the DDBJ website's search and analysis interface. At the top, there is a navigation bar with links for 'About DDBJ', 'How to Use', 'Report/Statistics', 'FAQ', and 'Contact Us'. Below this is a search bar with the text 'Google™ カスタム検索'. The main content area is titled 'Search and Analysis' and is divided into two columns. The left column is titled 'Database Search' and contains links for 'getentry' (Data retrieval by accession numbers, etc.), 'ARSA' (All-round Retrieval of Sequence and Annotation), 'TXSearch' (Retrieval of unified taxonomy database), 'BLAST' (Homology Search), 'VecScreen' (Vector Search to screen contamination in nucleic acid sequences), and 'DRA Search' (Search metadata by keywords and retrieve data). The right column is titled 'Genome Analyses' and contains links for 'MiGAP' (Mechanical annotation tool for microbial genomes), 'MiGAP-OLD (Result before Feb. 2012. Read-only)' (To login, DBCLS OpenID is required.), 'GTPS' (Reannotation of bacterial genomes using a new common), 'GTOP' (Genome to protein structure and function), and 'Links for DBCLS\* Search Tools'. Below these columns, there is a section titled 'Phylogenetics' which contains a link for 'ClustalW' (Multiple alignment and Tree-making).

**Database Search**

- [getentry](#)  
Data retrieval by accession numbers, etc.
- [ARSA](#)  
All-round Retrieval of Sequence and Annotation
- [TXSearch](#)  
Retrieval of unified taxonomy database
- [BLAST](#)  
Homology Search
- [VecScreen](#)  
Vector Search to screen contamination in nucleic acid sequences.
- [DRA Search](#)  
Search metadata by keywords and retrieve data

**Genome Analyses**

- [MiGAP](#)  
Mechanical annotation tool for microbial genomes(Login required.)
- [MiGAP-OLD \(Result before Feb. 2012. Read-only\)](#)  
To login, DBCLS OpenID is required.
- [GTPS](#)  
Reannotation of bacterial genomes using a new common
- [GTOP](#)  
Genome to protein structure and function
- [Links for DBCLS\\* Search Tools](#)

**Phylogenetics**

- [ClustalW](#)  
Multiple alignment and Tree-making

### Release of genome sequence data of *Oryza sativa* Japonica Group

<http://www.ddbj.nig.ac.jp/whatsnew/wn151013-e.html>

DDBJ released genome sequence data of *Oryza sativa* Japonica Group cv. Nipponbare Sequencing Project which had been submitted by National Institute of Agrobiological Sciences.

#### Reference

Database: RAP-DB

The accession numbers are as follows. They are available on [getentry](#).

#### chromosome:

AP014957-AP014968 (12 entries)

#### unanchored:

AP014969-AP015011 (43 entries)

HOME > 検索・解析 > getentry

getentry

### アクセッション番号等によるエントリ検索

ID : AP014957

①番号入力

DNA データベース: ☒ DDBJ ☐ GenBank ☐ EMBL ☐ MGA

Protein データベース: ☐ UniProt ☐ PDB ☐ DAD ☐ Patent

取得方法:

出力形式:

出力形式:

上限:  件

検索

②検索実行

```

LOCUS      AP014957              43270923 bp    DNA    linear    PLN 10-OCT-20
DEFINITION Oryza sativa Japonica Group DNA, chromosome 1, cultivar:
            Nipponbare, complete sequence.
ACCESSION  AP014957
VERSION    AP014957.1
DBLINK     BioProject:PRJDB1747
            BioSample:SAMD00000397
KEYWORDS   .
SOURCE     Oryza sativa Japonica Group (Japanese rice)
ORGANISM   Oryza sativa Japonica Group
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BOP
            clade; Oryzoideae; Oryzeae; Oryzinae; Oryza.
REFERENCE  1 (bases 1 to 43270923)
AUTHORS    Sakai,H., Kawahara,Y., Matsumoto,T., Buell,C.R. and Itoh,T.
TITLE      Direct Submission
JOURNAL     Submitted (09-OCT-2015) to the DDBJ/EMBL/GenBank databases.
            Contact:Takeshi Itoh
            National Institute of Agrobiological Sciences; Kannondai 2-1-2,
            Tsukuba 305-8602, Japan
REFERENCE  2
AUTHORS    Kawahara,Y., de la Bastide,M., Hamilton,J.P., Kanamori,H.,
            McCombie,W.R., Quyang,S., Schwartz,D.C., Tanaka,T., Wu,J.,
            Zhou,S., Childs,K.L., Davidson,R.M., Lin,H., Quesada-Ocampo,L.,
            Vaillancourt,B., Sakai,H., Lee,S.S., Kim,J., Numa,H., Itoh,T.,
            Buell,C.R. and Matsumoto,T.
TITLE      Improvement of the Oryza sativa Nipponbare reference genome using
            next generation sequence and optical map data.
JOURNAL     Rice 6, 4 (2013)
REFERENCE  3
AUTHORS    Sakai,H., Lee,S.S., Tanaka,T., Numa,H., Kim,J., Kawahara,Y.,
            Wakimoto,H., Yang,C.C., Iwamoto,M., Abe,T., Yamada,Y., Muto,A.,
            Inokuchi,H., Ikemura,T., Matsumoto,T., Sasaki,T. and Itoh,T.
TITLE      Rice Annotation Project Database (RAP-DB): an integrative and
            interactive database for rice genomics.
JOURNAL     Plant Cell Physiol. 54, e8 (2013)
COMMENT
FEATURES   Location/Qualifiers
            source          1..43270923
                        /chromosome="1"
                        /cultivar="Nipponbare"
                        /db_xref="taxon:39947"
                        /mol_type="genomic DNA"
                        /organism="Oryza sativa Japonica Group"
            assembly_gap    1..1000
                        /estimated_length=1000
                        /gap_type="telomere"
            CDS              join(3448..3616,4357..4455,5457..5560,7136..7944,
                        8028..8150,8232..8320,8408..8608,9210..9815,10102..10187,
                        10274..10297)
                        /codon_start=1
                        /gene="Os01t0100100"
                        /locus_tag="OSNPB_010100100"
                        /note="Os01t0100100-01: RabGAP/TBC domain containing
                        protein."
                        /protein_id="BAS89908.1"
                        /transl_table=1
                        /translation="MSSAAGGDNGDTAGDYIKWVGAGGRAGGAMANLQGVGSLY
                        IGDPCLNPSPVKGSKMLKPEKWHCTFDNDGKVIIGFRKALKFIVLGGVDPTRAEV
                        LLGCYALSSSTSEYRKLRAVRREKYQILVRQCGSMHPSIGTGELAYAVGSKLMDVF
                        SKETHIAEFVSTSQDTAGSIWFDSDYCPGGAGQSGKPEKSKSAELVGEN"

```

HOME > 検索・解析 > ARSA

### ARSA (Search Condition)

#### Quick Search

matK ①keyword  
Search AND  
②search button  
Available Fields

### ARSA (Search Result)

#### Search Condition

#### Search Result

##### Facet

##### List of Entries

1 - 60 entries / Number of founds: 124352 FlatFile XML Fasta View selected (10 entries) Download selected (10 entries)  
Download All

| PrimaryAccessionNumber                       | Definition  | SequenceLength       | MolecularType      | Organism                       |
|--|---|----------------------|--------------------|--------------------------------|
| <input checked="" type="checkbox"/> AJ519450 | Definition: Eltroplectris calcarata partial matK pseudogene for MatK protein. | SequenceLength: 1812 |                    |                                |
|  | MolecularType: DNA Organism: Eltroplectris calcarata                          |                      |                    |                                |
| <input checked="" type="checkbox"/> JX438642 | Definition: Ruppia maritima maturase K (matK) gene, partial cds; chloroplast. | SequenceLength: 799  |                    |                                |
|  | MolecularType: DNA Organism: Ruppia maritima                                  |                      |                    |                                |
| <input checked="" type="checkbox"/>          | :K) gene, partial cds; chloroplast gene for chloroplast product.              |                      |                    |                                |
| <input checked="" type="checkbox"/>          | ism: Cycas multipinnata   |                      |                    |                                |
| <input checked="" type="checkbox"/>          | ) gene, partial cds; chloroplast gene for chloroplast product.                |                      |                    |                                |
| <input checked="" type="checkbox"/>          | ism: Cycas taiwaniana   |                      |                    |                                |
| <input checked="" type="checkbox"/> AF410157 | Definition: Cycas fairylakea MatK (matK) gene, partial cds; chloroplast.      | SequenceLength: 1567 | MolecularType: DNA | Organism: Cycas fairylakea     |
| <input checked="" type="checkbox"/> AF410158 | Definition: Cycas panzhihuaensis MatK (matK) gene, partial cds; chloroplast.  | SequenceLength: 1567 | MolecularType: DNA | Organism: Cycas panzhihuaensis |
| <input checked="" type="checkbox"/> AF410160 | Definition: Cycas bifida MatK (matK) gene, partial cds; chloroplast.          | SequenceLength: 1567 | MolecularType: DNA | Organism: Cycas bifida         |
| <input checked="" type="checkbox"/> AF410159 | Definition: Cycas taitungensis MatK (matK) gene, partial cds; chloroplast.    | SequenceLength: 1567 | MolecularType: DNA | Organism: Cycas taitungensis   |
| <input checked="" type="checkbox"/> AF410161 | Definition: Cycas rumphii MatK (matK) gene, partial cds; chloroplast.         | SequenceLength: 1561 | MolecularType: DNA | Organism: Cycas rumphii        |
| <input checked="" type="checkbox"/> AF410163 | Definition: Cycas wadei MatK (matK) gene, partial cds; chloroplast.           | SequenceLength: 1561 | MolecularType: DNA | Organism: Cycas wadei          |
| <input type="checkbox"/> AF410162            | Definition: Cycas media MatK (matK) gene, partial cds; chloroplast.           | SequenceLength: 1561 | MolecularType: DNA | Organism: Cycas media          |
| <input type="checkbox"/> AF410164            | Definition: Cycas circinalis MatK (matK) gene, partial cds; chloroplast.      | SequenceLength: 1561 | MolecularType: DNA | Organism: Cycas circinalis     |
| <input type="checkbox"/> AF410166            | Definition: Dionon mejiae MatK (matK) gene, partial cds; chloroplast.         |                      |                    |                                |

④View  
(Fasta format)

③specify entries

⑤sequences

##### Facet

| Division  | Organism_facet  |
|---|---|
| <ul style="list-style-type: none"> <li>PLN(124143)</li> <li>ENV(42)</li> <li>PAT(41)</li> <li>EST(33)</li> <li>UNA(31)</li> <li>HUM(18)</li> <li>CON(17)</li> <li>SYN(9)</li> <li>TSA(5)</li> <li>GSS(4)</li> <li>BCT(3)</li> <li>ROD(2)</li> <li>HTC(1)</li> <li>MAM(1)</li> <li>STS(1)</li> <li>VRT(1)</li> </ul> | <ul style="list-style-type: none"> <li>Rheum tanguticum(305)</li> <li>Rheum palmatum(212)</li> <li>Orinus thordii(180)</li> <li>Brachypodium sylvaticum var. breviglume(154)</li> <li>Fallopia multiflora(141)</li> <li>Erica scoparia(136)</li> <li>Linaria elegans(120)</li> <li>Phoenix dactylifera(117)</li> <li>Armeria pungens(116)</li> <li>Erica arborea(105)</li> <li>Silene latifolia(104)</li> <li>Davallia repens(98)</li> <li>Centaurea maculosa subsp. maculosa(96)</li> <li>Quercus phillyraeoides(95)</li> <li>Capsicum chinense(91)</li> <li>Laurus nobilis(87)</li> </ul> |

# <clustalW>

## Generating phylogenetic tree

HOME > 検索・解析 > ClustalW

### ClustalW

#### Version

- ☒ 2.1 (Latest version)
- ☐ 1.83 (Modified by Dr. Kirill Kryukov)

#### Sequences

DNA sequences

File Upload: ファイルを選択 選択されていません  
or COPY & PASTE:

```
>AJ519450|AJ519450.1 Eltroplectris calcarata partial matK pseudogene for MatK protein.
cataacacacagaagtgcctccctttttgtttccagtaaaaaatgtatgcaatggcatta
ttacaagaatatttgaattgaaaaaaatcacatttctcaacaaaatttctatatccgc
tactcttcagg;
tttacgaacctg;
gtttaattactc;
taagaaggttt;
cagaaggttt;
aagaaaaaaagaccaaaatctcagaatttacatctattcatatcaatatttctttt
tagaagagaatttctcacattcaaatctgtgtcagatctactaataccatcccatcc
```

① Paste matK sequences

#### Submit

Send to ClustalW Clear

② Apply the button

Download Bootstrapped Tree File

③ Copy the result of Bootstrapped Tree

```
{
AF410161|AF410161.1_Cycas_rump:0.00125,
AF410163|AF410163.1_Cycas_wade:0.00067)
466:0.00043,
AJ519450|AJ519450.1_Eltroplect:0.11070,
JX438642|JX438642.1_Rupia_mar:0.11833)
1000:0.22045)
367:0.00027,
AF410155|AF410155.1_Cycas_mult:0.00205)
450:0.00045,
(AF410156|AF410156.1_Cycas_taiw:0.00000,
AF410157|AF410157.1_Cycas_fair:0.00000)
858:0.00108)
453:0.00020,
AF410160|AF410160.1_Cycas_bifi:0.00072,
(AF410158|AF410158.1_Cycas_panz:0.00011,
```

Home Download Gallery Docs Support ETE-tools<sup>beta</sup> Treeview

<http://et toolkit.org/treeview/>

A Python Environment

## Phylogenetic tree (newick) viewer

This is an online tool for phylogenetic tree view (newick format) that allows multiple sequence alignment (fasta format) to be shown together with the trees (fasta format). It uses the tree drawing engine implemented in the ETE toolkit. It offers transparent integration with the NCBI taxonomy database. Currently, alignments can be shown in condensed or block-based format. Leaf names in the newick tree should match those in the fasta format.

Tip: Use NCBI taxids as leaf names to get on-the-fly translation of species names and lineages.

Paste your tree in newick format:

```
AF410158|AF410158.1_Cycas_panz:0.00011,
AF410159|AF410159.1_Cycas_taiw:0.00117)
600:0.00056)TRICHOTOMY;
```

④ Paste the result of Bootstrapped Tree

tree view

⑤ Apply the button

Paste your alignment in fasta format:

block

```
4.5e+02 3.7e+02 4.7e+02 AF410161|AF410161.1_Cycas_rump
8.6e+02 AF410163|AF410163.1_Cycas_wade
1e+03 AJ519450|AJ519450.1_Eltroplect
JX438642|JX438642.1_Rupia_mar
4.5e+02 AF410155|AF410155.1_Cycas_mult
8.6e+02 AF410156|AF410156.1_Cycas_taiw
6e+02 AF410157|AF410157.1_Cycas_fair
AF410160|AF410160.1_Cycas_bifi
AF410158|AF410158.1_Cycas_panz
AF410159|AF410159.1_Cycas_taiw
```