Introduction of DNA Data Bank of Japan activities

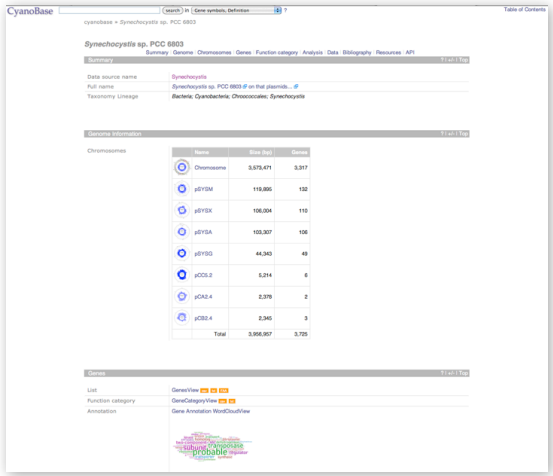
Genome Informatics Laboratory
National Institute of Genetics, JAPAN
Yaz Nakamura

Who am I?

I worked for Kazusa DNA Research Institute (1996-2008).
Analyses and DB construction for plant and plant-related bacteria genomes.
From 2009, I work for DDBJ, NIG.

The Arabidopsis Genome Initiative (2000)
Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 408, 796-815.

My team annotated 1/4 (27 Mb, 6200 genes) regions of *Arabidopsis* genome.



http://genome.microbedb.jp/cyanobase/
http://genome.microbedb.jp/rhizobase/

World central genome DB's for Cyanobacteria and Rhizobia (plant-related bacteria)

Genome Informatics Laboratory

Yaz Nakamura, Eli Kaminuma, Hideki Nagasaki, Takako Mochizuki, Takatomo Fujisawa, Naoko Iida, Yasuhiro Tanizawa, Naoko Murakata, Naoko Sakamoto

Genome informatics helps your life

DDBJ DNA Data Bank of Japan

GOLD Genomes Online Database

http://www.genomesonline.org/

JGI DOE JOINT GENOME INSTITUTE US DEPARTMENT OF ENERGY OFFICE OF SCIENCE

Genomes Online Database | Home | Version 4.0

Last update: 2013-05-15
Total # of genomes: **25618**
Download GOLD:

- Home
- Genome Map
- Genome Earth
- Search
- News
- Statistics
- Team
- Reference
- Contact

Blogger

Welcome to the Genomes OnLine Database

GOLD:Genomes Online Database, is a World Wide Web resource for comprehensive access to information regarding genome and metagenome sequencing projects, and their associated metadata, around the world.

Metagenomes

Classification

- Studies: **376**
- Samples: **2749**

Isolate Genomes

Complete Projects: **4331**

Incomplete Projects: **21210**

Targeted Projects: **1770**

Genome Distribution

- Project Type
- Sequencing Status
- Phylogenetic

1. Register

GOLD Genomes Online Database

Register your project information and Metadata in Genomes Online Database

Register

2. Annotate

Annotate your microbial genome or metagenome with IMG/ER or IMG/MER

Annotate

3. Publish

SIGS Standards in Genomic Sciences
An Open Access Journal of the Genomic Standards Consortium

Publish your genome or metagenome in open access standards-supportive journal.

Publish

U.S. DEPARTMENT OF ENERGY | Office of Science

My Lab's projects

CyanoBacteria, Rhizobia and Streptomyces DB's and TogoAnnotation

with DBCLS and Onishi lab. at Tokyo Univ. as an activity of integrated DB project

Citrus species' sequencing project

with National Institute of Fruit Tree Science and Fujiyama lab as a part of TRIC project

A liverwort (a moss) sequencing project

with Kohchi lab at Kyoto Univ. as a part of Genome Science project

A *Charophyceae* (an algae) annotation project

with Ohta lab at Titech

A Rubber tree sequencing project

with Bridgestone co ltd.

DNApod : DNA Polymorphism annOtation Database

as an activity of TRIC project

Supported by Grant-in-aids "Genome Science" by MEXT, Integrated Lifescience Database Project by JST and Transdisciplinary Research Integration Center (TRIC) project by ROIS

DDBJ

Osamu Ogasawara, Jun Mashima, Yuichi Kodama, Eli Kaminuma, Yasukazu Nakamura, Kousaku Okubo and Toshihisa Takagi (2013) DDBJ new system and service refactoring. *Nucl. Acids Res.,* **41 (D1):** D25-D29.

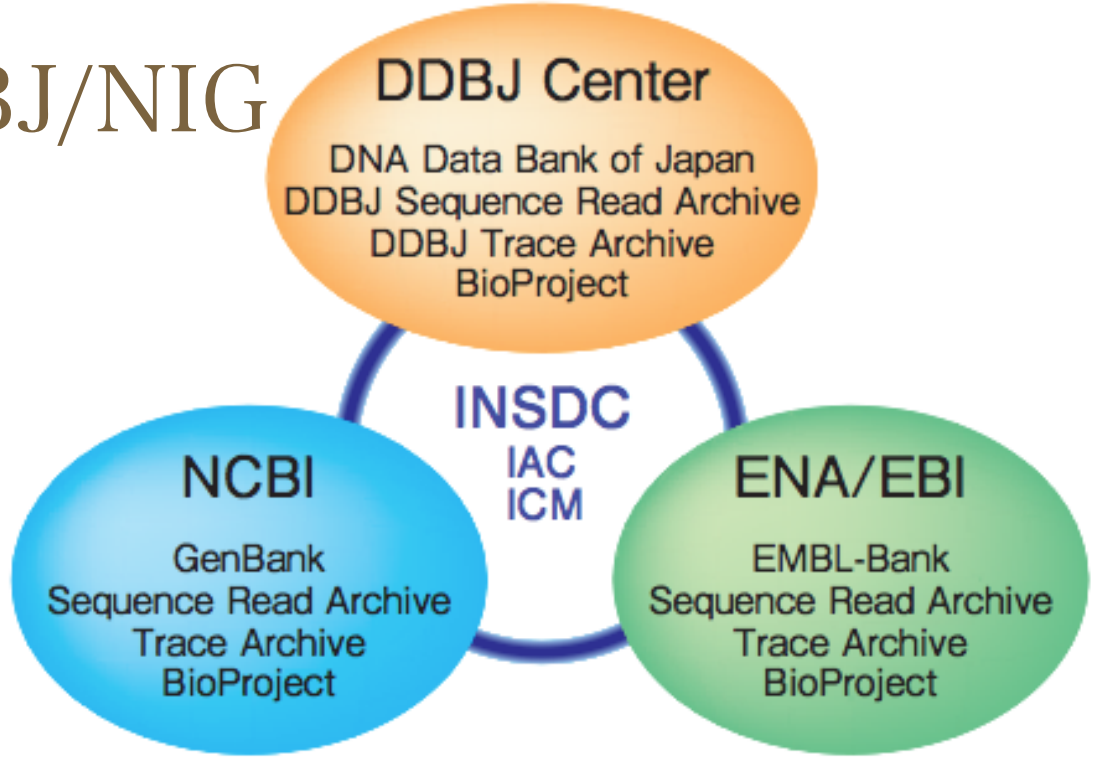
INSDC

Nakamura Y, Cochrane G, Karsch-Mizrachi I on behalf of the International Nucleotide Sequence Database Collaboration. (2013) The International Nucleotide Sequence Database Collaboration. *Nucleic Acids Res.,* **41 (D1),** D21-D24. Epub 2012 Nov 24.

DDBJ is a member of INSDC

International Nucleotide Sequence Databank Collaboration

- USA: GenBank/NCBI
- EU: ENA/EBI
- Japan: DDBJ/NIG



IAC: International Advisory Comittee
ICM: International Collaborative Meeting

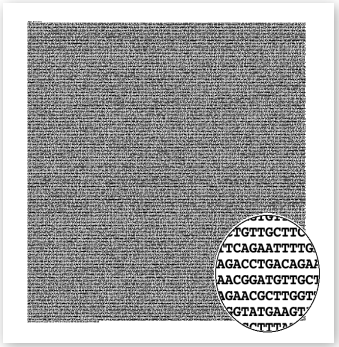
DDBJ (from Release note 92)
Jun Mashima, Hideo Aono, Yuji Ashizawa, Yukino Dobashi, Mayumi Ejima, Masahiro Fujimoto, Asami Fukuda, Tomohiro Hirai, Fumie Hirata, Naofumi Ishikawa, Toshikazu Katsumata, Chiharu Kawagoe, Shingo Kawahara, Yuichi Kodama, Junko Kohira, Takehide Kosuge, Kyungbum Lee, Mika Maki, Kimiko Mimura, Takeshi Moriyama, Yoshihisa Munakata, Naoko Murakata, Keiichi Nagai, Toshihisa Okido, Yoshihiro Okuda, Katsunaga Sakai, Makoto Sato, Yoshihiro Serizawa, Aimi Shiida, Yukie Shinyama, Rie Sugita, Kimiko Suzuki, Daisuke Takagi, Daisuke Takai, Haru Tsutsui, Koji Watanabe, Tomohiko Yasuda, Shigeru Yatsuzuka, Emi Yokoyama, Eli Kaminuma, Osamu Ogasawara, Kosaku Okubo, Toshihisa Takagi, and Yasukazu Nakamura

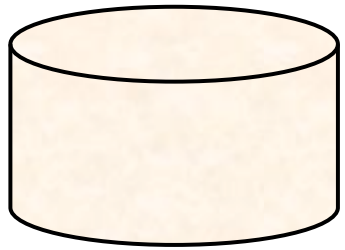
ENA (from Release note 115)
Blaise Alako, Clara Amid, Lawrence Bower, Ana Cerdeno-Taraga, Iain Cleland, Richard Gibson, Neil Goodgame, Petra ten Hoopen, Mikyung Jang, Simon Kay, Rasko Leinonen, Xin Liu, Arnaud Oisel, Rodrigo Lopez, Hamish McWilliam, Nima Pakseresht, Sheila Plaister, Rajesh Radhakrishnan, Kethy Reddy, Stephane Riviere, Marc Rossello, Nicole Silvester, Dmitriy Smirnov, Ana Luisa Toribio, Daniel Vaughan, Vadim Zalunin and Guy Cochrane

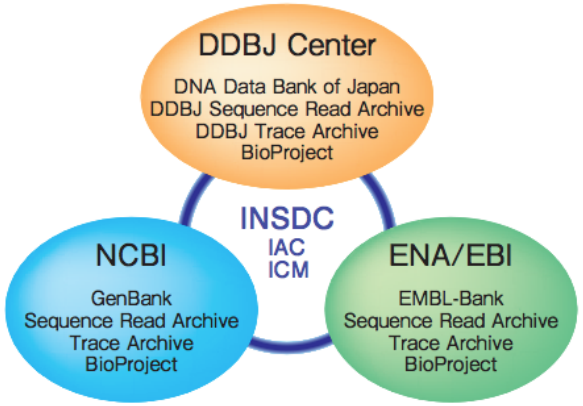
GenBank (from Release note 195)
Mark Cavanaugh, Ilene Mizrachi, Yiming Bao, Michael Baxter, Lori Black, Larissa Brown, Vincent Calhoun, Larry Chlumsky, Karen Clark, Jianli Dai, Michel Eschenbrenner, Irene Fang, Michael Fetchko, Linda Frisse, Andrea Gocke, Anjanette Johnston, Mark Landree, Jason Lowry, Suzanne Mate, Richard McVeigh, DeAnne Olsen Cravaritis, Leigh Riley, Susan Schafer, Beverly Underwood, Melissa Wright, Linda Yankie, Serge Bazhin, Evgueni Belyi, Colleen Bollin, Mark Cavanaugh, Yoon Choi, Ilya Dondoshansky, J. Bradley Holmes, WonHee Jang, Jonathan Kans, Leonid Khotomliansky, Michael Kimelman, Michael Kornbluh, Jim Ostell, Denis Sinyakov, Karl Sirotkin, Vladimir Soussov, Elena Starchenko, Hanzhen Sun, Tatiana Tatusova, Lukas Wagner, Eugene Yaschenko, Sergey Zhdanov, Slava Khotomliansky, Igor Lozitskiy, Craig Oakley, Eugene Semenov, Ben Slade, Constantin Vasilyev, Peter Cooper, Hanguan Liu, Wayne Matten, Scott McGinnis, Rana Morris, Steve Pechous, Monica Romiti, Eric Sayers, Tao Tao, Majda Valjavec-Gratian and David Lipman

The business of DNA Databank

- Determined Nucleotide Sequence
  - Checking Data and Metadata
  - Putting it into the Database
  - Open and Share it via the Internet







An example of DDBJ's entry

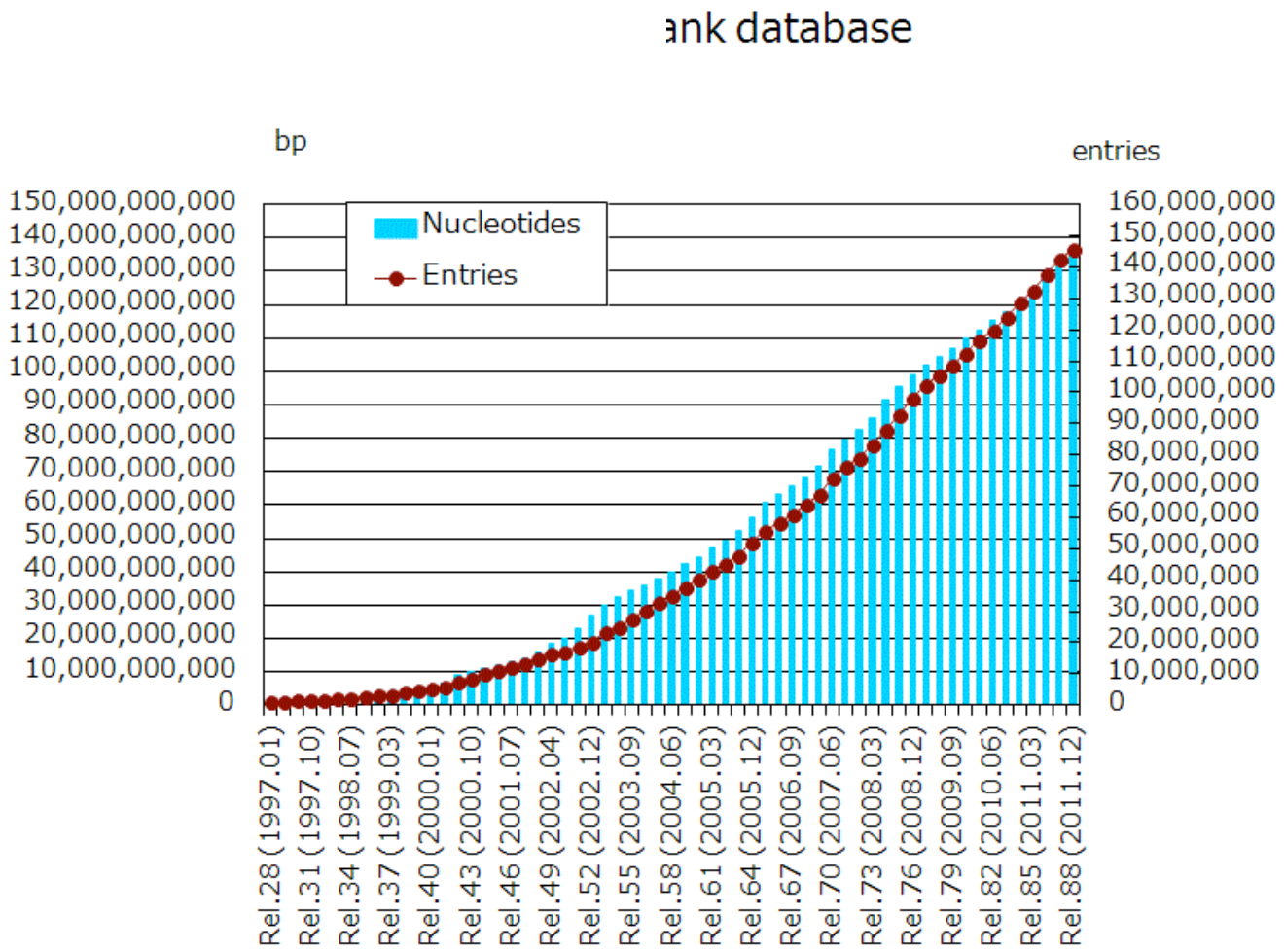
```
LOCUS       HUMIL2HOM                397 bp    DNA     linear   HUM 27-APR-1993
DEFINITION  Human interleukin 2 (IL-2)-like DNA.
ACCESSION   M13784
VERSION     M13784.1
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 397)
  AUTHORS   Mita,S., Maeda,S. and Shimada,K.
  TITLE     Characterization of human genomic DNA sequences homologous to the
            interleukin 2 cDNA
  JOURNAL   Biochem. Biophys. Res. Commun. 138 (2), 966-973 (1986)
   PUBMED   3017347
COMMENT     Original source text: Human placenta DNA, clone Lm HoIL2-3.
            Numerous stop codons are found in the interleukin 2-like IIa DNA.
FEATURES             Location/Qualifiers
     source          1..397
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT          117 a           84 c           48 g          148 t
ORIGIN      RsaI site.
        1 actgatttat ttttaataaa attacaagag attttaattt taaacccaaa agttctttta
       61 ttgcatctca ctgtgtttag ctttgtttac cctttgagaa ggcctgagat aataactttc
      121 ttcttcaact ctttcatcag ctcctgtaac ctttttttcct taggttctta actgatgttg
      181 tggcctgctg ctaaaaacgc tttatcttaa agttctaaaa ggaaatgttt tcttctaaca
      241 taacattctg ggctcttgac tttatgaaat caaaaacttt cacttatgac caggatacac
      301 tcttcctctg tctaactaat tcaagcacta tcttcattca ttttgacttg cagattatcc
      361 aaacagactc cccataatga aaagcaatca cactgca
//
```

Current data amount of INSDC

DDBJ DNA Data Bank of Japan

140 billion nucleotides　　150 million entries

1,400億　　1.5億



Note: CON division is not counted in statistics of DDBJ periodical

Species' word cloud in INSDC

DDBJ
DNA Data Bank of Japan

The cloud gives greater prominence to words that appear more frequently in the source text.

INSDC 2013 International Advisory Committee meeting

Guy Cochrane

14 May 2013:
12:00-15:00 in UK
07:00-09:30 in USA (DC)
20:00-22:30 in Japan
(teleconference)

EMBL-EBI

Summary of technical agenda

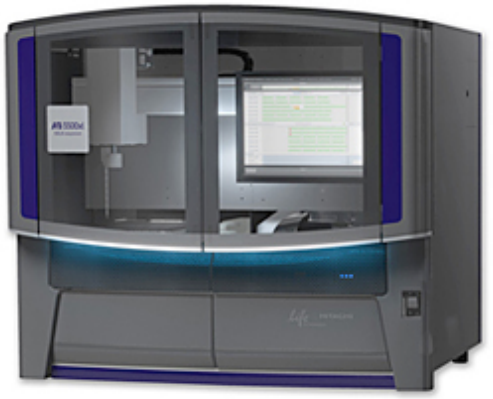
- Meeting to be held next week over three days
- 17 core participants
- 9 sessions
  - Feature table
  - **Third Party Annotation**
  - **Genomes and assemblies**
  - **Contigs and sets**
  - Samples
  - Taxonomy
  - **Feature table evolution**
  - Data models
  - Patents and INSDseq

21-23 May 2013: Cambridge, UK

DB’s and Services for

NGS

example of the NGS's

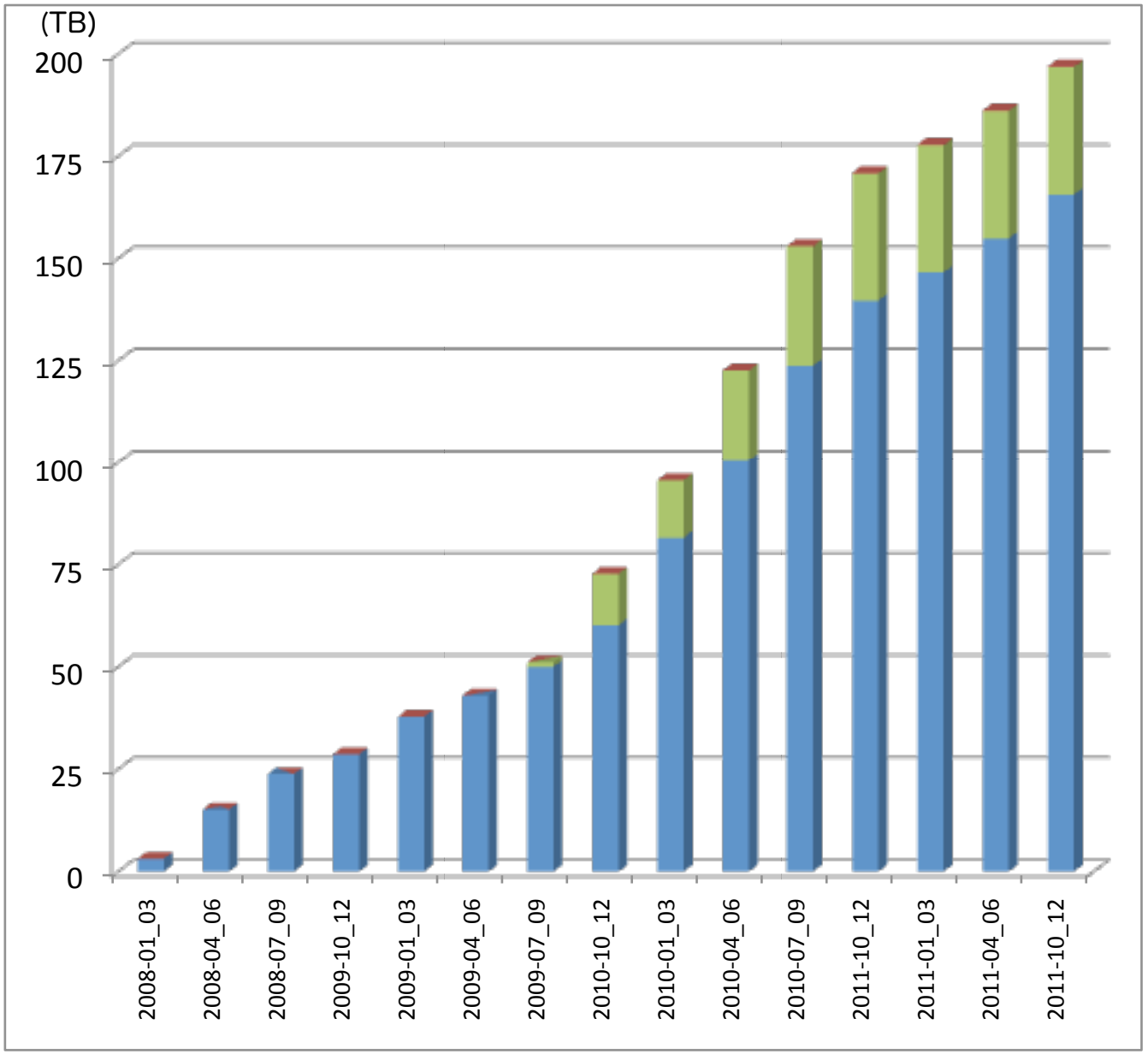


Roche (454): GS FLX+ System

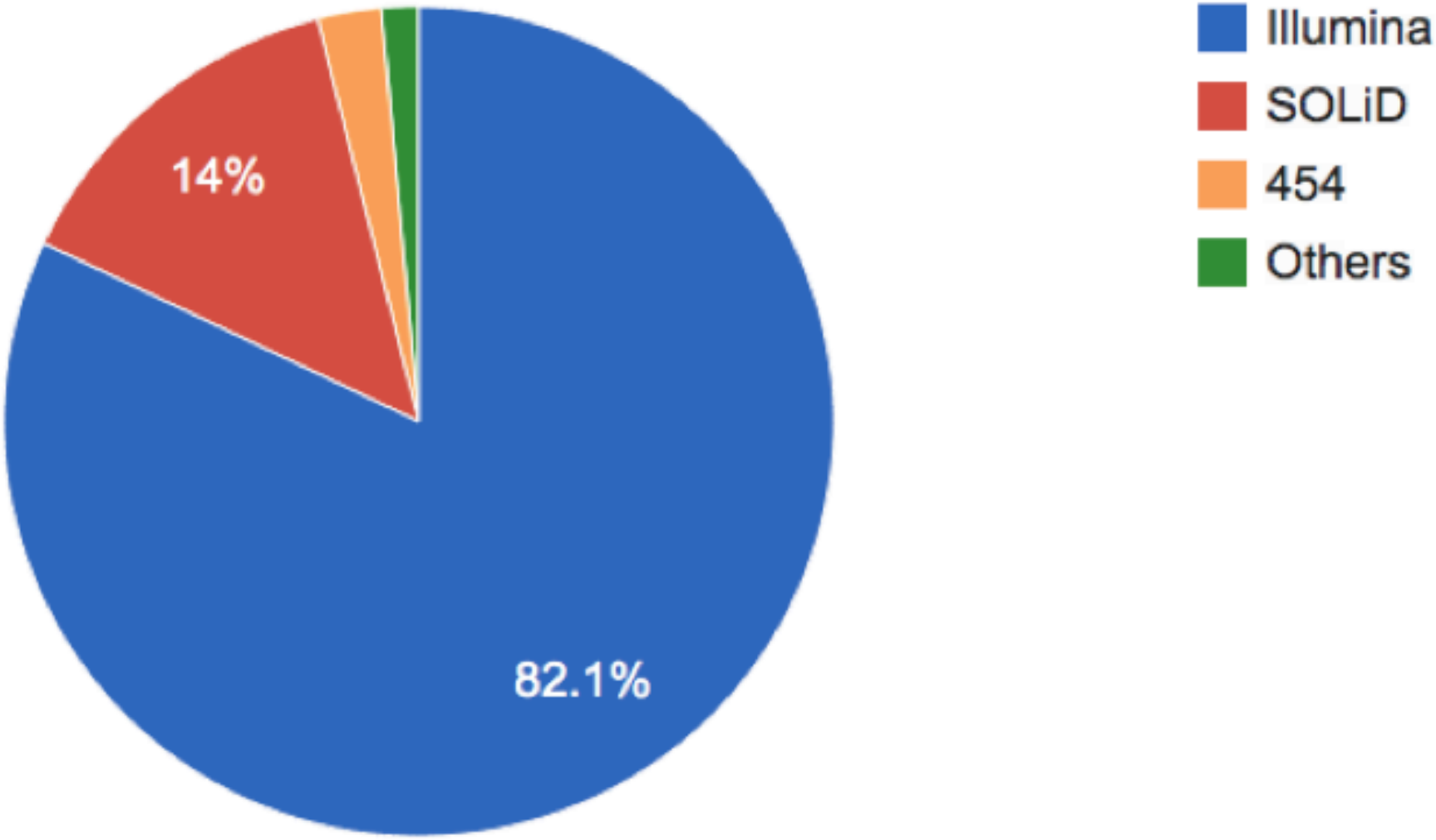
illumina: Genome Analyzer IIx System

Life Technologies: 5500 xl SOLiD System

SRA's data amount: a problem in Japan

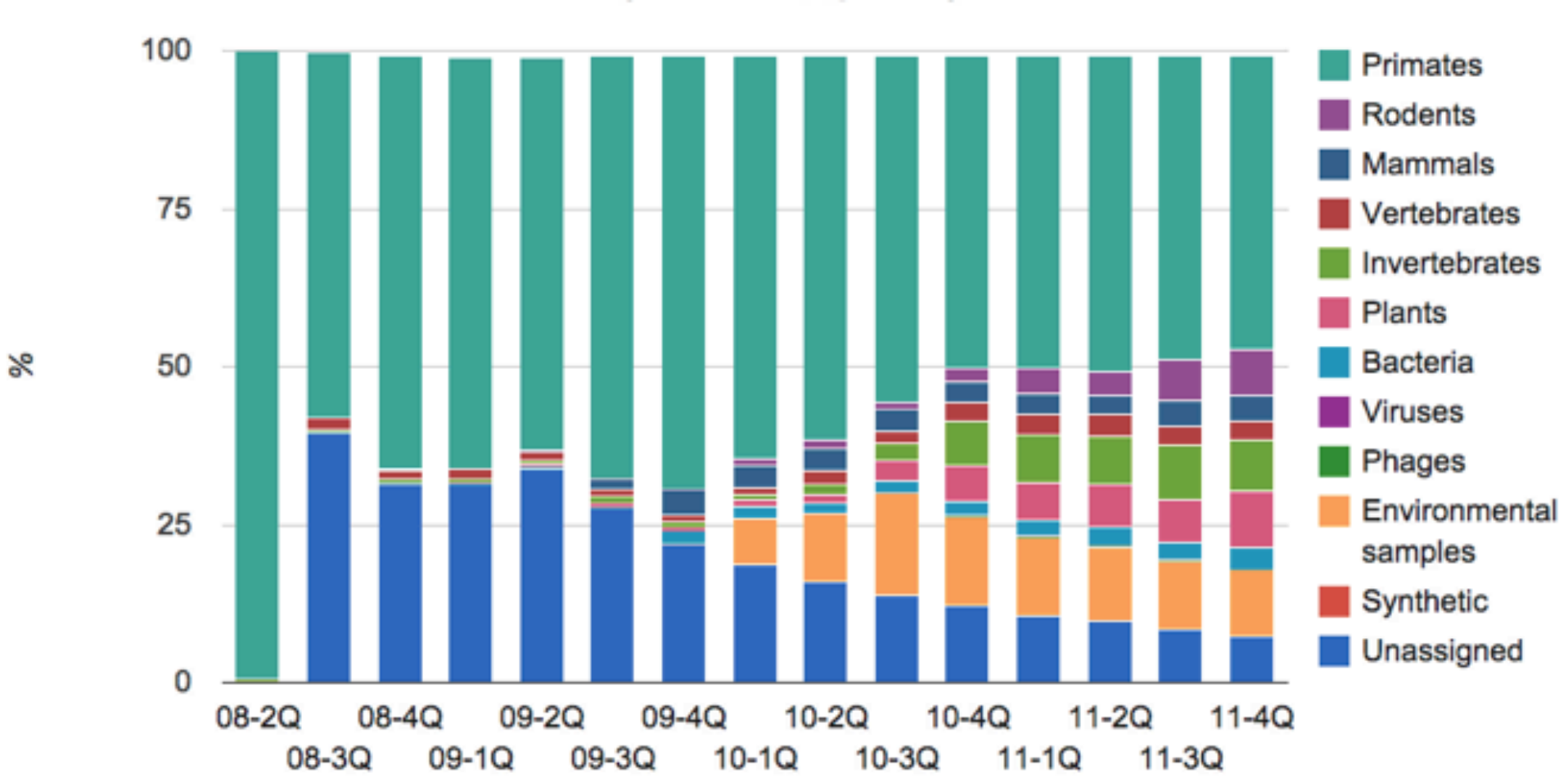


Sequencing platform in SRA (in TB)



http://trace.ddbj.nig.ac.jp/dra/stat_sra.shtml

Taxonomic division in SRA (in TB)



http://trace.ddbj.nig.ac.jp/dra/stat_sra.shtml

DDBJ Sequence Read Archive (DRA)

http://trace.ddbj.nig.ac.jp/dra/

DRA: NGS data's archive

DDBJ
DNA Data Bank of Japan

DDBJ Sequence Read Archive

DDBJ Sequence Read Archive | DDBJ Trace Archive | DDBJ BioProject

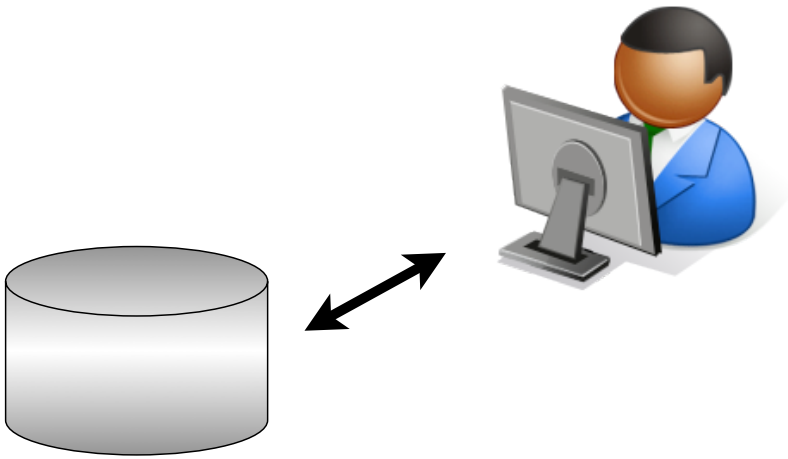
Home | Documentation | Submission | Search | Download | Pipeline | About

DDBJ Sequence Read Archive (DRA) は Roche 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD®
ークエンサからの出力データのためのデータベースです。DRA は International Nucleotide Sequence Database Collaboration (IN
り， NCBI Sequence Read Archive (SRA) と EBI Sequence Read Archive (ERA) との国際協力のもと，運営されています。従来
ンサからの出力データは DDBJ Trace Archive にご登録ください。

登録に必要なメタデータとデータファイル
登録方法
データの検索・ダウンロード
DDBJ Read Annotation Pipeline でデータを解析

» 動画マニュアル
MetaDefine 動画マニュアル
データ転送動画マニュアル

Site Policy | Privacy | © DNA Data Bank of Japan



Released Data

search result : 30

Accession	Study Title	Organism(s)	Center Name	Release Date
DRA000001	Whole genome sequencing of Baillus subtilis subsp. natto BEST195	Bacillus subtilis subsp. natto	KEIO	2010-03-26
DRA000002	Whole genome resequencing of Bacillus subtilis subsp. subtilis str. 168	Bacillus subtilis subsp. subtilis str. 168	KEIO	2010-03-26
DRA000010	Whole genome shotgun sequences of Oryza sativa japonica variety, Koshihikari	Oryza sativa Japonica Group	NIAS	2010-03-31
DRA000030	Whole-genome DNA methylation analysis in human breast cancer cell lines using MeDIP-seq	Homo sapiens	KUGSPS	2010-03-01
DRA000039	genetic variation detected in 206 klebsiella pneumoniae plasmids	Klebsiella pneumoniae	WMC	2009-12-14
DRA000067	B. anthracis BA103 genome analysis	Bacillus anthracis	NIID	2010-04-22
DRA000068	B. anthracis BA104 genome analysis	Bacillus anthracis	NIID	2010-04-22
DRA000069	Whole SNPs analysis of ciprofloxacin resistance among B. anthracis strains	Bacillus anthracis	NIID	2010-04-22
DRA000070	Whole SNPs analysis of ciprofloxacin resistance among B. anthracis strains	Bacillus anthracis	NIID	2010-04-22
DRA000155	CAGE analysis of whole adult brain and whole embryo rat transcriptome	Rattus norvegicus	RIKEN_OSC	2010-03-17
DRA000169	Linking new promoters to functional transcripts in small samples with nanoCAGE and CAGEscan	Homo Sapiens	RIKEN_OSC	2010-06-08
DRA000205	A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness	Homo Sapiens	RIKEN_OSC	2010-07-23
DRA000220	Whole genome sequencing of Oryzias latipes Hd-rR	Oryzias latipes	KEIO-SM	2010-08-16
SRA002052	Toxoplasma gondii transcript sequencing project	Toxoplasma gondii	UT-MGS	2009-07-01
SRA002053	Glossina morsitans transcript sequencing project	Glossina morsitans	UT-MGS	2009-07-01
SRA002054	Glossina morsitans transcript sequencing project	Glossina morsitans	UT-MGS	2009-06-25
SRA002055	Anopheles stephensi transcript sequencing project	Anopheles stephensi	UT-MGS	2009-07-01
SRA002056	Cryptosporidium parvum transcript sequencing project	Cryptosporidium parvum	UT-MGS	2009-07-01
SRA002057	Plasmodium yoelii transcript sequencing project	Plasmodium yoelii	UT-MGS	2009-09-22

FTP ディレクトリ /ddbj_database/dra/ / ftp.ddbj.nig.ac.jp

この FTP サイトはエクスプローラーでは表示するには、ページ をクリックして、エクスプローラーで FTP サイトを開く をクリックしてください。

```
-Welcome to DDBJ FTP Archive, running on ftp.ddbj.nig.ac.jp!
-
-Please contact ddbj@ddbj.nig.ac.jp when you have any problem for getting
-access to this archive, downloading the data, and etc.
-
-Termination of DDBJ-XML output format.
-Here is the announcement.
-http://www.ddbj.nig.ac.jp/whatsnew/2010/100625-e.html
-
-A new directory, "patent", was made under "ddbj_database".
-Now, all of patent amino acid sequence data for JPO and KIPO are
-included in the new "patent" directory.
-
-For details, please read the README.TXT in this directory.
-
-Distributions of the latest DDBJ release and newly-arrived/updated
-entries after that release can be retrieved at the following FTP
-sites.
-
-DDBJ Flat file
-ftp://ftp.ddbj.nig.ac.jp/ddbj_database/ddbj/
-ftp://ftp.ddbj.nig.ac.jp/ddbj_database/ddbjnew/
-
-DDBJ Flat file(TPA)
-ftp://ftp.ddbj.nig.ac.jp/ddbj_database/tpa/ddbj/
-ftp://ftp.ddbj.nig.ac.jp/ddbj_database/tpa/ddbjnew/
-
-INSD-XML
-ftp://ftp.ddbj.nig.ac.jp/ddbj_database/ddbj/xml/insdxml_current/
-ftp://ftp.ddbj.nig.ac.jp/ddbj_database/ddbjnew/xml/insdxml_current/
-
-It was last modified on Fri Feb 26 2010.
```

↑階層上のディレクトリへ

```
07/30/2010 08:07午後      ディレクトリ DRA000
07/05/2010 01:27午後      ディレクトリ ERA000
07/30/2010 08:08午後      ディレクトリ ERA005
07/30/2010 08:09午後      ディレクトリ ERA007
07/30/2010 08:09午後      ディレクトリ ERA008
07/30/2010 08:09午後      ディレクトリ ERA009
07/30/2010 08:08午後      ディレクトリ ERA010
07/13/2010 04:54午前      ディレクトリ SRA000
07/13/2010 09:26午前      ディレクトリ SRA001
```

DRA's new services (DRAsearch)

DDBJ DNA Data Bank of Japan

DRASearch Send Feedback | Search Home | DRA Home

Accession :

Organism :
CenterName :

StudyType :
Platform :

Keyword :

Show 20 records Sort by Study Search Clear

- ✓
- 16S rRNA tag sequencing
- Cancer Genomics
- copy number variation
- Deep Amplicon Sequencing
- environmental amplicon sequencing
- Epigenetics
- Exome Sequencing
- Forensic or Paleo-genomics
- Functional Genomics
- Gene Regulation Study
- genome partial sequencing
- Genome Variation Profiling
- Genomics
- Metagenomics
- multi-isolate
- Other
- Paleo-genomics
- Paleo-genomics array capture
- pooled clone
- Pooled Clone Sequencing
- Population Genomics
- RAD Sequencing
- Resequencing
- RIP-CHIP
- RNASeq
- small RNA
- sorted chromosome sequencing
- subtractive hybridization
- Synthetic Genomics
- Taq ssu rRNA gene hypervariable region taq sequencing
- Transcriptome Analysis
- Whole genome bisulfite sequencing
- Whole Genome Sequencing

Data Last Update 2012-02-10
WebSite Last Update 2011-06-20

Statistics

Released Entries

Type	Count
Submission	60744
Study	9509
Experiment	119718
Sample	252390
Run	380019

Organism

#	Organism Name	Study
1	unidentified	875
2	Homo sapiens	812
3	Mus musculus	447
4	Drosophila melanogaster	207
5	metagenome sequence	179
6	Caenorhabditis elegans	143
7	marine metagenome	141
8	Escherichia coli str. K-12 substr. MG1655	133
9	Mustela putorius furo	100
10	Arabidopsis thaliana	98

#		
1	Whole G	
2	Transcr	
3	Metage	
4	Epigene	
5	Resequencing	498
6	RNASeq	335
7	Other	281
8	Population Genomics	145
9	Gene Regulation Study	51
10	Exome Sequencing	44

ne [All List]

#	lame	Study
		1554
		1326
		1298
		868
5	UMIGS	818
6	WUGSC	380
7	SC	347
8	BCM	296
9	NCBI	70
10	CFSAN	59

An user's voice (tweets)

Oh! Year! (me)

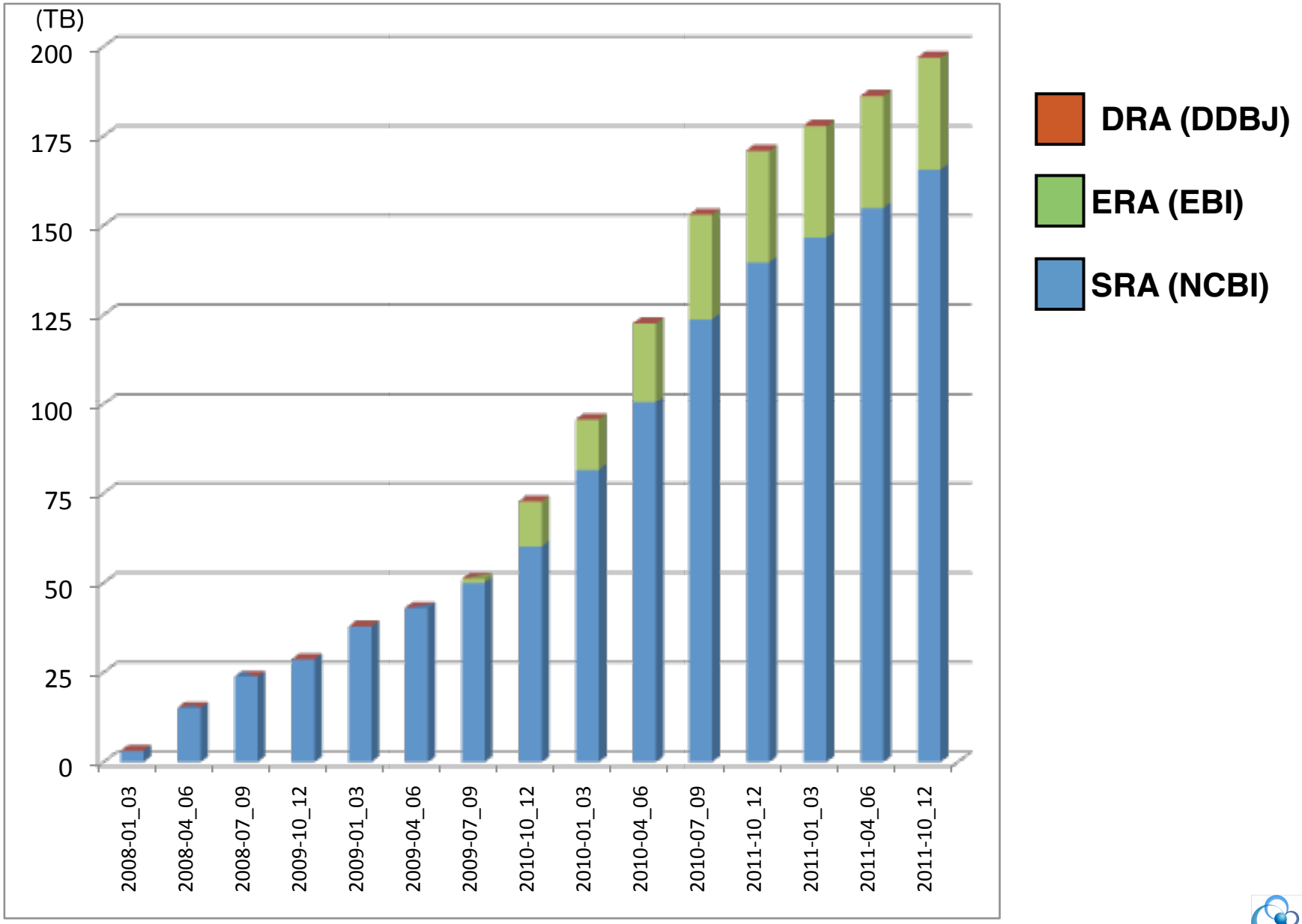
Not nessessary to use SRA site.

DDBJ DRA is fast to download!

Easy to understand!

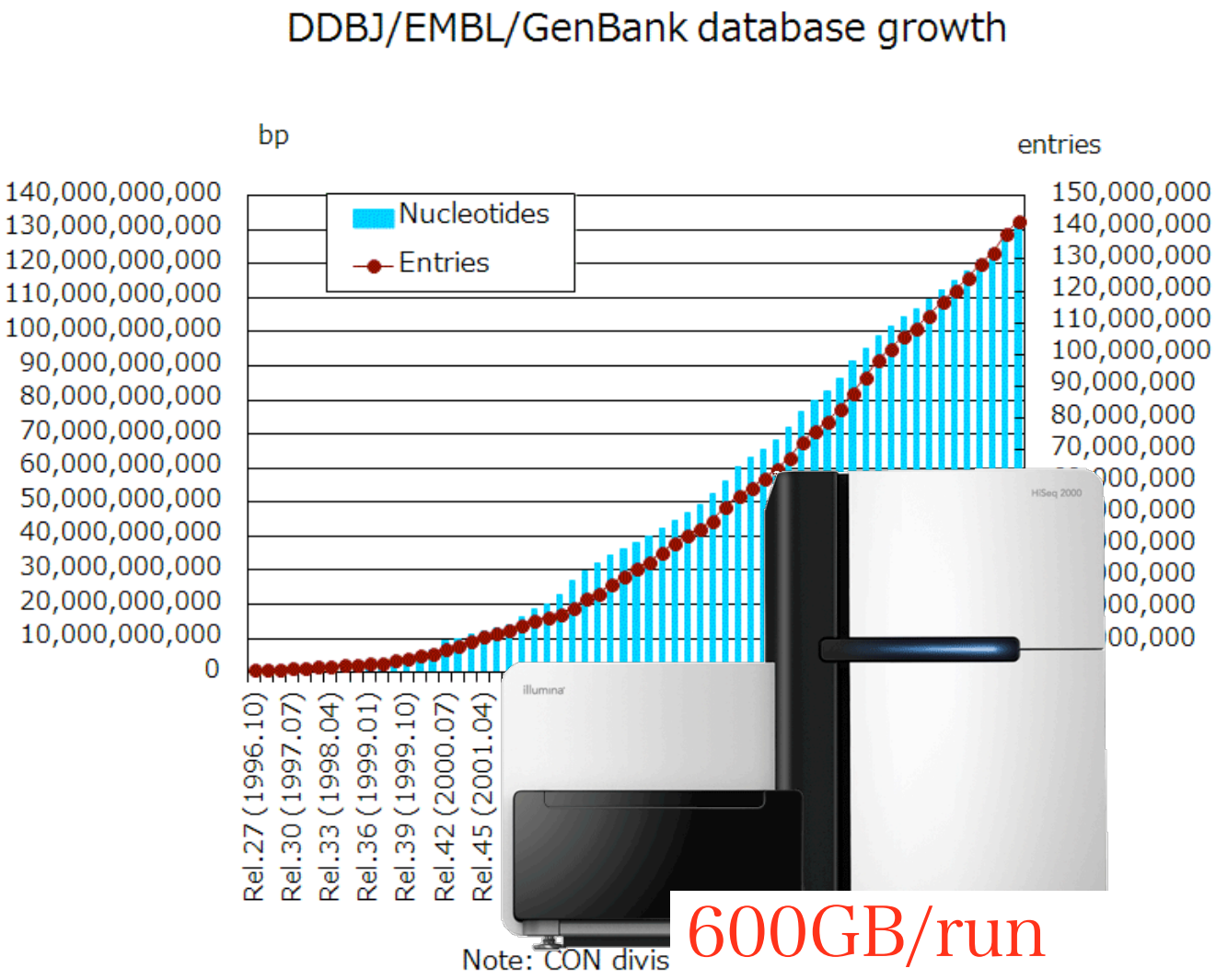
DRAsearch is extremely handy!

DRA's data amount: a problem in Japan



Data explosion (again)

Trad DB: 140GB



“NANOPORE” sequencer will be emerged

http://www.nanoporetech.com/technology/introduction-to-nanopore-sensing/introduction-to-nanopore-sensing

search contact us

Oxford NANOPORE Technologies®

Home | Technology | About Us | News | Careers

Technology

- Introduction to nanopore sensing
  - Biological nanopores
  - Solid state nanopores
- The GridION system
- Single use cartridge
- GridION workflow
- GridION informatics
- MinION: a miniaturised sensing instrument
- Analytes and Applications: DNA, RNA, proteins
- Fields of use
- Publications

Introduction to nanopore sensing

The concept of using a nanopore as a biosensor was first proposed in the mid 1990s when nanopores were starting to be researched at academic institutions such as Oxford, Harvard and UCSC - all Oxford Nanopore collaborators. In an industrial setting, Oxford Nanopore was founded in 2005 to translate nanopore science into an electronics-based technology. The end-to-end system includes sample preparation, molecular analysis and informatics, and is designed to provide disruptive user benefits in a number of applications.

Oxford Nanopore has a broad intellectual property portfolio that includes internal innovation and collaborations with world leading nanopore researchers. This IP includes fundamental nanopore sensing techniques through to solid-state nanopore sensing technology including graphene.

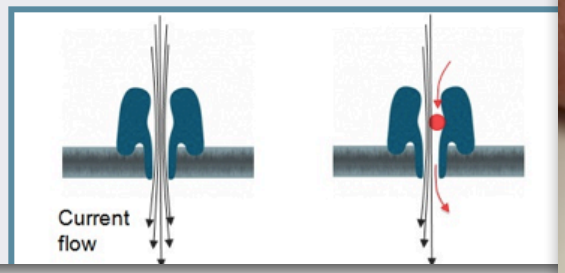
Nanopore fabrication

A nanopore is, essentially, a nano-scale hole. This hole may be:

- Biological: formed by a pore-forming protein in a membrane such as a lipid bilayer
- Solid-state: formed in synthetic materials such as silicon nitride or g
- Hybrid: formed by a pore-forming protein set in synthetic material

Nanopore sensing

A nanopore may be used to identify a target analyte as follows.



“MinION - $900 usb-powered DNA sequencer”

We are faced with “Big Data”

DDBJ
DNA Data Bank of Japan

data

Strange things in the Refseq / nr DB

DDBJ
DNA Data Bank of Japan

- protain (183) < protein
- imilar to (28) < similar to
- simila to (22) < similar to
- cromosome (4) < chromosome
- RNA olymerase < RNA polymerase
- dehydrogenas, ehydrogenase
- transposas, ransposase
- “2-Sep” for septin-2 < SEPT2

Copy & paste error!

DDBJ
DNA Data Bank of Japan

>gi|91204169|emb|CAJ71822.1| strongly **imilar to** aspartate aminotransferase [Candidatus Kuenenia stuttgartiensis]
MIASRMSNIDSSGIRKVFDLAQKMKSPVNLSIGQPDFDVPGEIKEVAIKSINEGANKYTLTQGIPELRNV
...

>gi|31541577|gb|AAP56877.1| predicted methyl **transferas** [Mycoplasma gallisepticum R]
MSALYLVGLPIGNLSEINHRALEILNQLEIIYCENTDNFKKLLNLLNINFRDKKLISYHKFNETNRFIMI
...

similar to
transferase

“similar to similar to”

```
LOCUS       AL591981             347050 bp    DNA     linear   BCT 16-APR-2005
DEFINITION  Listeria monocytogenes strain EGD, complete genome, segment 9/12.
ACCESSION   AL591981 AL591824
VERSION     AL591981.1
KEYWORDS    .
SOURCE      Listeria monocytogenes
  ORGANISM  Listeria monocytogenes
            Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE   2  (bases 1 to 347050)
  AUTHORS   Glaser,P., Frangeul,L. and Rusniok,C.
  JOURNAL   Submitted (06-JUN-2001) to the EMBL/GenBank/DDBJ databases. Glaser
            P., Institut Pasteur, Genomique des Microorganismes Pathogenes, 25
            rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE.

...
     CDS             complement(12915..14294)
                     /transl_table=11
                     /gene="lmo1703"
                     /note="similar to similar to RNA methyltransferases"
                     /db_xref="GOA:Q8Y6I1"
                     /db_xref="InterPro:IPR001566"
                     /db_xref="InterPro:IPR002792"
                     /db_xref="InterPro:IPR010280"
                     /db_xref="UniProtKB/Swiss-Prot:Q8Y6I1"
                     /protein_id="CAC99781.1"
                     /translation="MNQNPVEEGQKFPLTIRRMGINGEGIGYFKKAVVFVPGAITGEEV
                     VVEAVKVRDRFTEAKLNKIRKKSPNRVTAPCPVYEACGGCQLQHVAYSAQLELKRDIVI
                     QSIEKHTKIDPTKLKIRPTIGMEDPWRYRNKSQFQTRMVGSGQVETGLFGANSHQLVPI
                     EDCIVQQPVTIKVTNFVRDLLEKYGVPIYDEKAGSGIVRTIVVRTGVKTGETQLVFITN
                     SKKLPKKREMLAEIEAALPEVTSIMQNVNQAKSSLIFGDETFLLAGKESIEEKLMELEF
                     DLSARAFFQLNPFQTERLYQEVEKALVLTGSETLVDAYCGVGTIGQAFAGKVKEVRGMD
                     IIPESIEDAKRNAEKNGIENVYYEVGKAEDVLPKWVKEGFRPDAVIVDPPRSGCDQGLI
                     KSLLDVEAKQLVYVSCNPSTLARDLALLAKKYRIRYMQPVDMFPQTAHVETVVLLQLKD
                     K"
```

SEPT2 ⇒ 2-Sep case in Refseq

Introduced by MS-Excel's automatic correction

```
LOCUS       XM_392412               2125 bp    mRNA    linear   INV 12-APR-2011
DEFINITION  PREDICTED: Apis mellifera septin-2 (2-Sep), mRNA.
ACCESSION   XM_392412
VERSION     XM_392412.4  GI:328785636
KEYWORDS    .
SOURCE      Apis mellifera (honey bee)
  ORGANISM  Apis mellifera
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
            Apidae; Apis.
COMMENT     MODEL REFSEQ:  This record is predicted by automated computational
            analysis. This record is derived from a genomic sequence
            (NW_003378075) annotated using gene prediction method: GNOMON,
            supported by EST evidence.
            Also see:
                Documentation of NCBI's Annotation Process

            On Apr 12, 2011 this sequence version replaced gi:110757583.
FEATURES             Location/Qualifiers
     source          1..2125
                     /organism="Apis mellifera"
                     /mol_type="mRNA"
                     /strain="DH4"
                     /db_xref="taxon:7460"
                     /linkage_group="LG6"
     gene            1..2125
                     /gene="2-Sep"
                     /note="Derived by automated computational analysis using
                     gene prediction method: GNOMON. Supporting evidence
                     includes similarity to: 436 ESTs, 11 Proteins"
                     /db_xref="BEEBASE:GB17411"
                     /db_xref="GeneID:408882"
     misc_feature    164..166
                     /gene="2-Sep"
                     /note="upstream in-frame stop codon"
     CDS             194..1444
                     /gene="2-Sep"
                     /codon_start=1
                     /product="septin-2"
                     /protein_id="XP_392412.2"
```

http://www.biomedcentral.com/1471-2105/5/80

Identifier “mutation” by Excel (2-Sep)

DDBJ DNA Data Bank of Japan

Correspondence

Mistaken Identifiers: Gene name errors can be introduced inadvertently when using Excel in bioinformatics

Barry R Zeeberg[1†], **Joseph Riss**[2†], **David W Kane**[3], **Kimberly J Bussey**[1], **Edward Uchio**[4], **W Marston Linehan**[4], **J Carl Barrett**[2] and **John N Weinstein**[1*]

* Corresponding author: John N Weinstein weinstein@dtpvx2.ncifcrf.gov
† Equal contributors

Author Affiliations

1 Genomics & Bioinformatics Group, Laboratory of Molecular Pharmacology, Center for Cancer Research (CCR), National Cancer Institute (NCI), National Institutes of Health (NIH), Bldg 37 Rm 5041, NIH, 9000 Rockville Pike, Bethesda, MD 20892 USA
2 Laboratory of Biosystems and Cancer, CCR, Bldg 37 Rm 5032, NIH, 9000 Rockville Pike, Bethesda, MD 20892 USA
3 SRA International, 4300 Fair Lakes CT, Fairfax, VA 22033 USA
4 Urologic Oncology Branch, Bldg 10 Rm 2B47, National Institutes of Health, Bethesda, MD 20892 USA

For all author emails, please log on.

BMC Bioinformatics 2004, **5**:80 doi:10.1186/1471-2105-5-80

How to Avoid such stupid Errors?

DDBJ
DNA Data Bank of Japan

- Good Reference Sequences
- Good Reference Annotations
- Fully Automated Annotation process
  - Ontology for gene and metadata
  - Rule-based gene description

- No copy and paste by hand
- No auto-correction by Excel

cloud

+

crowd

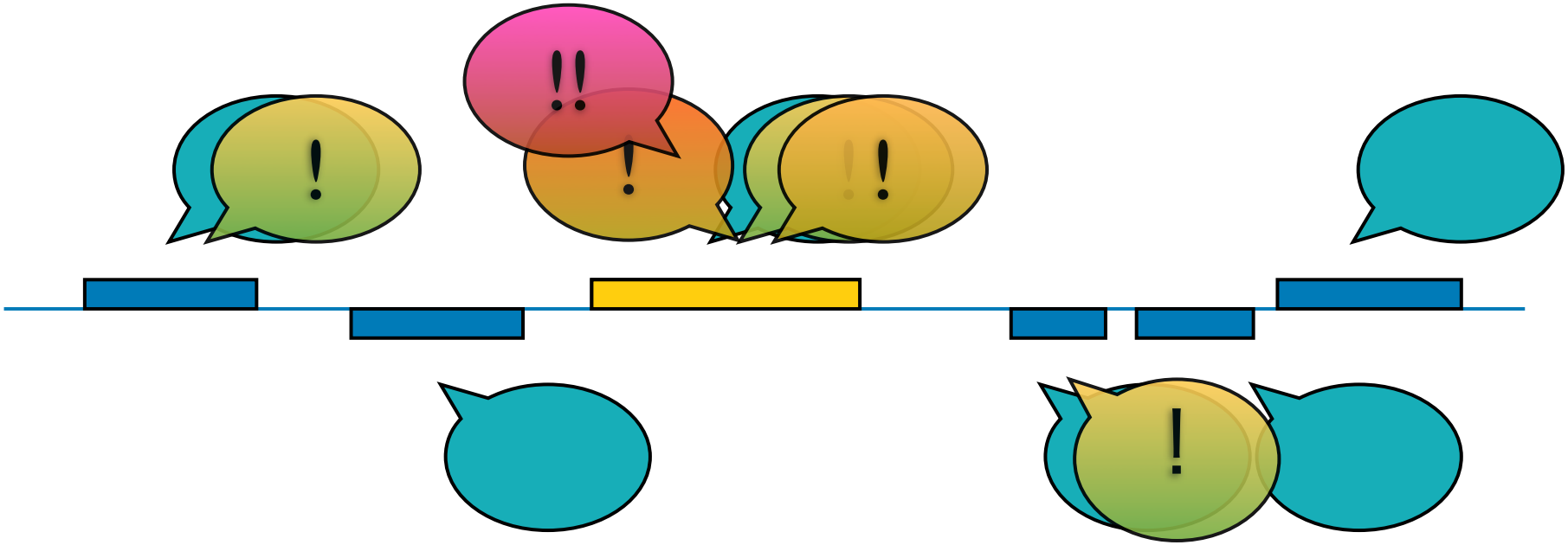
We need:

- good references
- good pipelines

We need:

- good references
- good pipelines

TogoAnnotation: a social genome annotation platform



http://togo.annotation.jp

http://togo.annotation.jp

TogoAnnotation

検索 もっと検索する...

サインイン

About TogoAnnotation

統合アノテーションは、ソーシャルブックマークのしくみを利用して生物のデータに様々な注釈（アノテーション）をつけることが出来るサービスです。

News

統合アノテーション
TogoAnnotation

TogoAnnotation TogoAnnotation (KazusaAnnotation)を用いてコミュニティゲノムアノテーションを実施したBradyrhizobium sp. S23321の論文、配列が公開されました。 jstage.jst.go.jp/article/jsme2/... getentry.ddbj.nig.ac.jp/getentry?acces...
25 days ago · reply · retweet · favorite

bonohu TogoAnnotationｷﾀ――――(∀)――――!! (#AJACS live at ustre.am/1x4x/1)
38 days ago · reply · retweet · favorite

synobu 19種の光合成関連生物から約8000報の文献をキュレーターが読んで20万以上の文献中の遺伝子名記述を抽出しています. togo.annotation.jp
40 days ago · reply · retweet · favorite

synobu TogoAnnotation: ソーシャルブックマークでゲノム注釈を行なうプロジェクト. 遺伝子, 文献, ゲノムごとのサマリや各種JSON APIが付け加わりました. togo.annotation.jp
40 days ago · reply · retweet · favorite

yaskaz KazusaAnntation 改め TogoAnnotation review 中。さらにかっちょ良くなったお togo.annotation.jp

Join the conversation

Recent Annotations

Ann:TG:4742
http://streptomyces.nih.go.jp/gview/view_annon.cgi?molecule=TG&id=4...
1405 annotations annotation:157874

Ann:TG:2095
http://streptomyces.nih.go.jp/gview/view_annon.cgi?molecule=TG&id=2...
60 annotations annotation:158993

Ann:TG:4742
http://streptomyces.nih.go.jp/gview/view_annon.cgi?molecule=TG&id=4...

Recent Genes

AdpA, adpAg
http://streptomyces.nih.go.jp/gview/view_annon.cgi?molecule=TG&id=4742
1405 annotations
26 references

sgmA
http://streptomyces.nih.go.jp/gview/view_annon.cgi?molecule=TG&id=2095
60 annotations
14 references

adsA
http://streptomyces.nih.go.jp/gview/view_annon.cgi?molecule=TG&id=4151
78 annotations
17 references

more... Last updated: 2012-04-24 04:09

Recent References

Scherzinger, D. Ruch, S. Kloer, D. P. Wilde, A. Al-Babili, S.
Retinal is formed from apo-carotenoids in Nostoc sp. PCC7120:...
Biochem J. 2006 Sep 15;398(3):361-9.
72 annotations pmid:16759173

Tomono, A. Tsai, Y. Ohnishi, Y. Horinouchi, S.
Three chymotrypsin genes are members of the AdpA regulon in...
J Bacteriol. 2005 Sep;187(18):6341-53.
346 annotations pmid:16159767

Zhao, K. H. Zhang, J. Tu, J. M. Bohm, S. Ploscher, M. Eichacker, L. Bubenzer, C. Sc...
Lyase activities of CpcS- and CpcT-like proteins from Nostoc PC...
J Biol Chem. 2007 Nov 23;282(47):34093-103. Epub 2007 Sep 25.
9 annotations pmid:17895251

more... Last updated: 2012-04-24 08:36

Recent Genomes

http://togo.annotation.jp

References ? | +/- | Top

CyanoBase search in Gene symbols, Definition ?

1. Shoumskaya, M. A. Paithoonrangsarid, K. Kanesaki, Y. Los, D. A. Zinchenko, V. V. Tanticharoen, M. Suzuki, I. Murata, N.
Identical Hik-Rre systems are involved in perception and transduction of salt signals and hyperosmotic signals but regulate the expression of individual genes to different extents in synechocystis.
J Biol Chem. 2005 Jun 3;280(22):21531-8. Epub 2005 Mar 31. PMID:15805106 | Abstract | MeSH | Related Articles | Sections | Gene Index |
Ti Ab In Mm Re Di Co Fi Ta | CyanoGenes:1712 CyanoGenes:MedlineID gname:rre31 GI abstract introduction experimental procedures results discussion table1 table2 fig1 fig2 fig3 fig4

2. Ashby, M. K. Mullineaux, C. W.
Cyanobacterial ycf27 gene products regulate energy transfer from phycobilisomes to photosystems I and II.
FEMS Microbiol Lett. 1999 Dec 15;181(2):253-60. PMID:10585546 | Abstract | MeSH | Related Articles | Sections | Gene Index |
Ti Ab In Mm Re Di Co Fi Ta | CYORF:CYREF GI syn:slr0115 abstract materials and methods results discussion table1 table2 fig1 fig2 fig3

3. Ashby, M. K. Houmard, J. Mullineaux, C. W.
The ycf27 genes from cyanobacteria and eukaryotic algae: distribution and implications for chloroplast evolution.
FEMS Microbiol Lett. 2002 Aug 27;214(1):25-30. PMID:12204368 | Abstract | MeSH | Related Articles | Sections | Gene Index |
Ti Ab In Mm Re Di Co Fi Ta | DBTCY GI tf_type:OmpR gname:rpaA tfac abstract introduction materials methods discussion table1 table3 table4 fig1

4. Paithoonrangsarid, K. Shoumskaya, M. A. Kanesaki, Y. Satoh, S. Tabata, S. Los, D. A. Zinchenko, V. V. Hayashi, H. Tanticharoen, M. Suzuki, I. Murata, N.
Five histidine kinases perceive osmotic stress and regulate distinct sets of genes in Synechocystis.
J Biol Chem. 2004 Dec 17;279(51):53078-86. Epub 2004 Oct 7. PMID:15471853 | Abstract | MeSH | Related Articles | Sections | Gene Index
Ti Ab In Mm Re Di Co Fi Ta | CyanoGenes:1712 CyanoGenes:MedlineID DBTCY GI tf_type:OmpR tfac abst results discussion table2 fig1 fig2 fig3

5. Hanke, G. T. Satomi, Y. Shinmura, K. Takao, T. Hase, T.
A screen for potential ferredoxin electron transfer partners uncovers new, redox dependent interactions.
Biochim Biophys Acta. 2011 Feb;1814(2):366-74. Epub 2010 Sep 22. PMID:20869472 | Abstract | MeSH | Related Articles | Sections | Gene Index
Ti Ab In Mm Re Di Co Fi Ta | GI gname:rpaA gname:rre31 abstract results discussion conclusions table3 fig2

6. Kappell, A. D. van Waasbergen, L. G.
The response regulator RpaB binds the high light regulatory 1 sequence upstream of the high-light-inducible hliB gene from the cyanobacterium Synechocystis PCC 6803.
Arch Microbiol. 2007 Apr;187(4):337-42. Epub 2007 Feb 10. PMID:17294172 | Abstract | MeSH | Related Articles | Sections | Gene Index
Ti Ab In Mm Re Di Co Fi | GI materials and methods results discussion fig2 fig3

7. Tabei, Y. Okada, K. Tsuzuki, M.
Sll1330 controls the expression of glycolytic genes in Synechocystis sp. PCC 6803.
Biochem Biophys Res Commun. 2007 Apr 20;355(4):1045-50. Epub 2007 Feb 22. PMID:17331473 | Abstract | MeSH | Related Articles | Sections | Gene Index |
Ti Ab In Mm Re Di Co Fi Ta | GI gname:rpaA results discussion table1 fig2

8. Sugita, C. Ogata, K. Shikata, M. Jikuya, H. Takano, J. Furumichi, M. Kanehisa, M. Omata, T. Sugiura, M. Sugita, M.
Complete nucleotide sequence of the freshwater unicellular cyanobacterium Synechococcus elongatus PCC 6301 chromosome: gene content and organization.
Photosynth Res. 2007 Jul-Sep;93(1-3):55-67. Epub 2007 Jan 9. PMID:17211581 | Abstract | MeSH | Related Articles | Sections | Gene Index |
Ti Ab In Mm Re Di Co Fi Ta | GI gname:rpaA gname:rre31 results discussion table3

9. Murata, N. Suzuki, I.
Exploitation of genomic sequences in a systematic analysis to access how cyanobacteria sense environmental stress.
J Exp Bot. 2006;57(2):235-47. Epub 2005 Nov 29. PMID:16317040 | Abstract | MeSH | Related Articles | Sections | Gene Index |
Ti Ab In Mm Re Di Co Fi Ta | CYORF:CYREF CyanoGenes:1712 CyanoGenes:MedlineID GI gname:rre31 syn:slr0115 body fig4

10. Mary, I. Vaulot, D.
Two-component systems in Prochlorococcus MED4: genomic analysis and differential expression under stress.
FEMS Microbiol Lett. 2003 Sep 12;226(1):135-44. PMID:13129619 | Abstract | MeSH | Related Articles | Sections | Gene Index |
Ti Ab In Mm Re Di Co Fi Ta | GI discussion table4

11. Takai, N. Nakajima, M. Oyama, T. Kito, R. Sugita, C. Sugita, M. Kondo, T. Iwasaki, H.
A KaiC-associating SasA-RpaA two-component regulatory system as a major circadian timing mediator in cyanobacteria.
Proc Natl Acad Sci U S A. 2006 Aug 8;103(32):12109-14. Epub 2006 Aug 1. PMID:16882723 | Abstract | MeSH | Related Articles | Sections | Gene Index |
Ti Ab In Mm Re Di Co Fi Ta | GI gname:rre31 gname:rpaA discussion

12. Mizuno, T. Kaneko, T. Tabata, S.
Compilation of all genes encoding bacterial two-component signal transducers in the genome of the cyanobacterium, Synechocystis sp. strain PCC 6803.
DNA Res. 1996 Dec 31;3(6):407-14. PMID:9097043 | Abstract | MeSH | Related Articles | Sections | Gene Index |
Ti Ab In Mm Re Di Co Fi Ta | DBTCY GI tf_type:OmpR tfac table2

13. Lechno-Yossef, S. Fan, Q. Ehira, S. Sato, N. Wolk, C. P.
Mutations in four regulatory genes have interrelated effects on heterocyst maturation in Anabaena sp. strain PCC 7120.
J Bacteriol. 2006 Nov;188(21):7387-95. Epub 2006 Aug 25. PMID:16936023 | Abstract | MeSH | Related Articles | Sections | Gene Index |
Ti Ab In Mm Re Di Co Fi Ta | GI gname:rre31 introduction

14. Hihara, Y. Kamei, A. Kanehisa, M. Kaplan, A. Ikeuchi, M.
DNA microarray analysis of cyanobacterial gene expression during acclimation to high light.
Plant Cell. 2001 Apr;13(4):793-806.

15. Midorikawa, T. Matsumoto, K. Narikawa, R. Ikeuchi, M.
An Rrf2-type transcriptional regulator is required for expression of psaAB genes in the cyanobacterium Synechocystis sp. PCC 6803.

16. Peter, ... Wallner, T. W... A. Grimm, B.

Manually curated,
<u>complete publication</u>
list for each gene

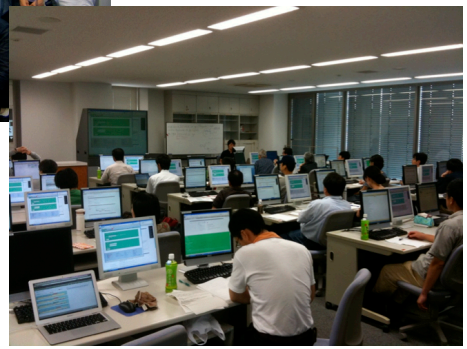
For example, 429 papers for a gene *psbA2* at:

http://genome.kazusa.or.jp/cyanobase/Synechocystis/genes/slr1311

TogoAnnotation: applications

- Gene indexing: paper curation project (>5000)
  - Cyanobacteria, Rhizobia, Streptmyces
- Community annotation projects.
  - Comminity annotation for a *Rhizobium*

3 days training
then
2 months
remote



Microbes Environ. 2012 Mar 28. [Epub ahead of print]

Complete Genome Sequence of *Bradyrhizobium* sp. S23321: Insights into Symbiosis Evolution in Soil Oligotrophs.

Okubo T, Tsukui T, Maita H, Okamoto S, Oshima K, Fujisawa T, Saito A, Futamata H, Hattori R, Shimomura Y, Haruta S, Morimoto S, Wang Y, Sakai Y, Hattori M, Aizawa SI, Nagashima KV, Masuda S, Hattori T, Yamashita A, Bao Z, Hayatsu M, Kajiya-Kanegae H, Yoshinaga I, Sakamoto K, Toyota K, Nakao M, Kohara M, Anda M, Niwa R, Jung-Hwan P, Sameshima-Saito R, Tokuda SI, Yamamoto S, Yamamoto S, Yokoyama T, Akutsu T, Nakamura Y, Nakahira-Yanaka Y, Takada Hoshino Y, Hirakawa H, Mitsui H, Terasawa K, Itakura M, Sato S, Ikeda-Ohtsubo W, Sakakura N, Kaminuma E, Minamisawa K.

- Currently we annotate an algae and a moss genomes.

http://togo.annotation.jp

TogoAnnotation

検索 もっと検索する...

サインイン

About TogoAnnotation

統合アノテーションは、ソーシャルブックマークのしくみを利用して生物のデータに様々な注釈（アノテーション）をつけることが出来るサービスです。

News

統合アノテーション
TogoAnnotation

TogoAnnotation TogoAnnotation (KazusaAnnotation)を用いてコミュニティゲノムアノテーションを実施したBradyrhizobium sp. S23321の論文、配列が公開されました。 jstage.jst.go.jp/article/jsme2/... getentry.ddbj.nig.ac.jp/getentry?acces...
25 days ago · reply · retweet · favorite

bonohu TogoAnnotationｷﾀ――――(∀)――――!! (#AJACS live at ustre.am/1x4x/1)
38 days ago · reply · retweet · favorite

synobu 19種の光合成関連生物から約8000報の文献をキュレーターが読んで20万以上の文献中の遺伝子名記述を抽出しています. togo.annotation.jp
40 days ago · reply · retweet · favorite

synobu TogoAnnotation: ソーシャルブックマークでゲノム注釈を行なうプロジェクト. 遺伝子, 文献, ゲノムごとのサマリや各種JSON APIが付け加わりました. togo.annotation.jp
40 days ago · reply · retweet · favorite

yaskaz KazusaAnntation 改め TogoAnnotation review 中。さらにかっちょ良くなったお togo.annotation.jp

Join the conversation

Recent Annotations

Ann:TG:4742
http://streptomyces.nih.go.jp/gview/view_annon.cgi?molecule=TG&id=4...
1405 annotations annotation:157874

Ann:TG:2095
http://streptomyces.nih.go.jp/gview/view_annon.cgi?molecule=TG&id=2...
60 annotations annotation:158993

Ann:TG:4742
http://streptomyces.nih.go.jp/gview/view_annon.cgi?molecule=TG&id=4...

Recent Genes

AdpA, adpAg
http://streptomyces.nih.go.jp/gview/view_annon.cgi?molecule=TG&id=4742
1405 annotations
26 references

sgmA
http://streptomyces.nih.go.jp/gview/view_annon.cgi?molecule=TG&id=2095
60 annotations
14 references

adsA
http://streptomyces.nih.go.jp/gview/view_annon.cgi?molecule=TG&id=4151
78 annotations
17 references

more... Last updated: 2012-04-24 04:09

Recent References

Scherzinger, D. Ruch, S. Kloer, D. P. Wilde, A. Al-Babili, S.
Retinal is formed from apo-carotenoids in Nostoc sp. PCC7120:...
Biochem J. 2006 Sep 15;398(3):361-9.
72 annotations pmid:16759173

Tomono, A. Tsai, Y. Ohnishi, Y. Horinouchi, S.
Three chymotrypsin genes are members of the AdpA regulon in...
J Bacteriol. 2005 Sep;187(18):6341-53.
346 annotations pmid:16159767

Zhao, K. H. Zhang, J. Tu, J. M. Bohm, S. Ploscher, M. Eichacker, L. Bubenzer, C. Sc...
Lyase activities of CpcS- and CpcT-like proteins from Nostoc PC...
J Biol Chem. 2007 Nov 23;282(47):34093-103. Epub 2007 Sep 25.
9 annotations pmid:17895251

more... Last updated: 2012-04-24 08:36

Recent Genomes

We need:

DDBJ
DNA Data Bank of Japan

- good references
- good pipelines

NIG Supercomputer (2012.3-)

2012.03.01

Phase 1

- 165.1 TFlops
- 5 PB HDD
- Containing 10TB and 2TB shared memory system.

Rmax of LINPACK: 82.90 TFLOPS

Rank:170th in Top500 (Nov. 2011)

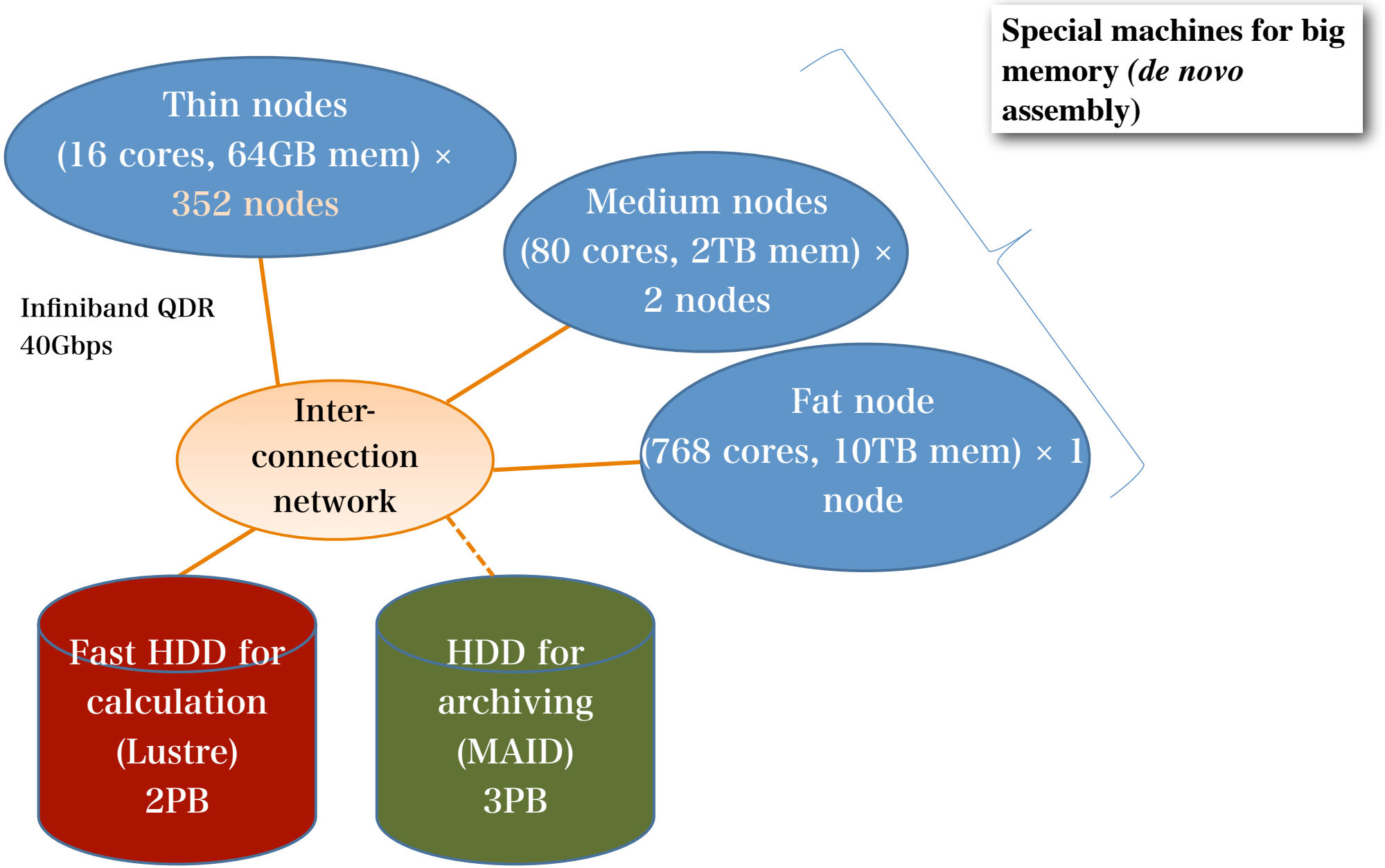
(Rank 11th in Japan)

2014.03.01

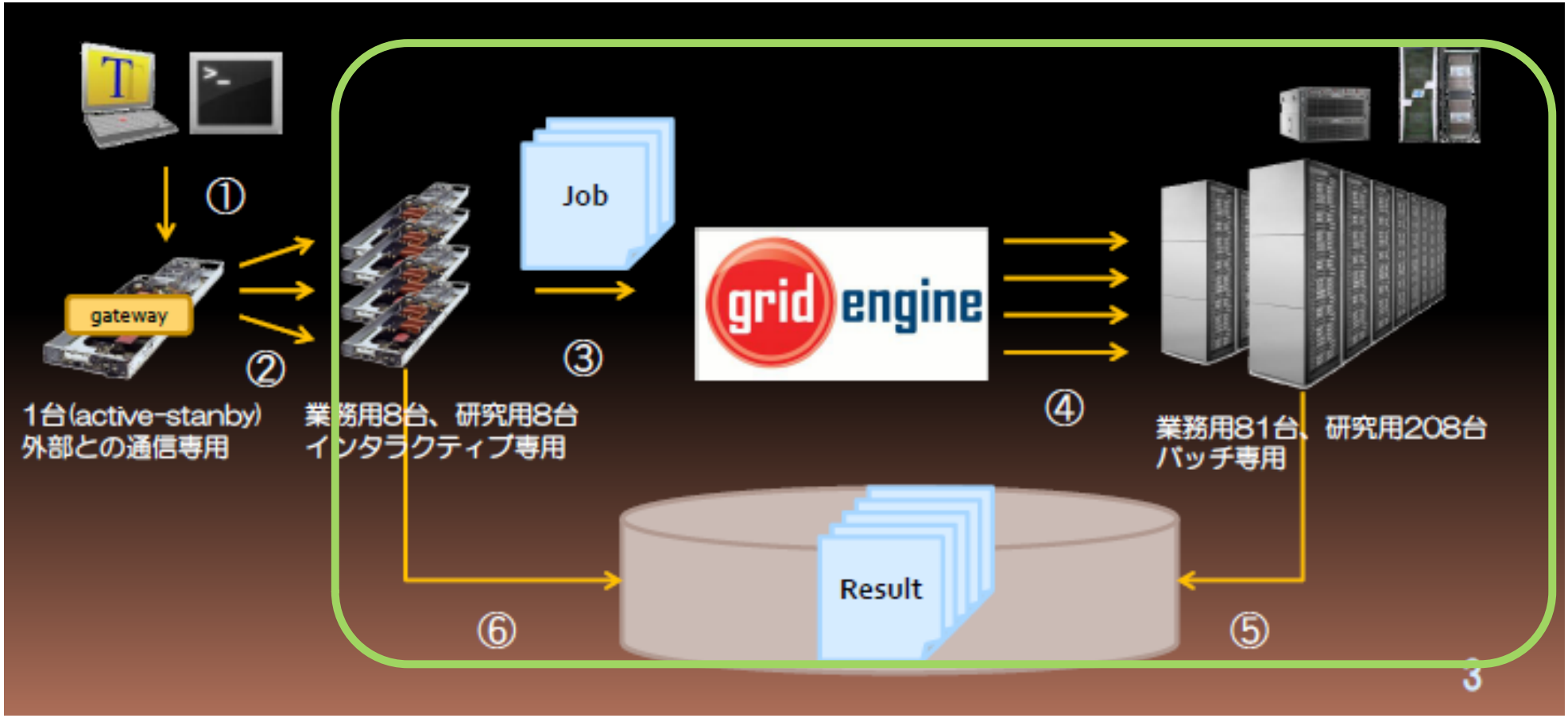
Phase 2

- about 400 TFlops (total)
- 12.5 PB HDD (total)

NIG Supercomputer overview



Running Batch Jobs



Running Jobs on Medium and Fat nodes

DDBJ
DNA Data Bank of Japan

```
# run a script on a thin node.
qsub -cwd -S /bin/bash your_script.sh

# run a script on a medium node.
qsub -cwd -l month -l medium -S /bin/bash
your_script.sh

# run a script on the fat node.
qsub -cwd -l month -l fat -S /bin/bash
your_script.sh
```

Memory Request (for each CPU core)

DDBJ
DNA Data Bank of Japan

```
# This job runs on 1 CPU core and 128GB
memory.
qsub -cwd -l month -l medium
 -l s_vmem=128G,mem_req=128G
 -S /bin/bash your_script.sh

# This job runs on 10 CPU core (in the same
node) and 1280GB memory.
qsub -cwd -l month -l medium
 -l s_vmem=128G,mem_req=128G
 -pe def_slot=10
 -S /bin/bash your_script.sh
```

It’s easy to use, isn’t it?

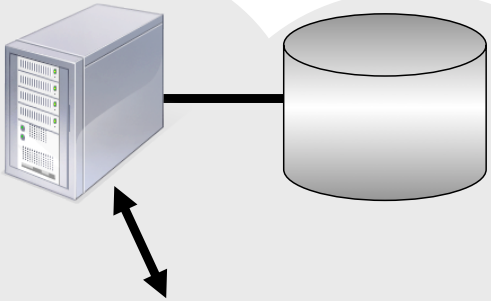
(簡単でしょ？)

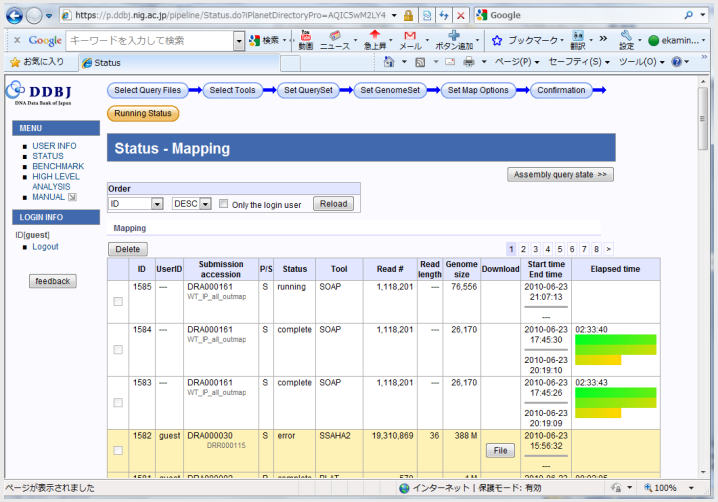
(゜д゜) ...What?

(ハァ？)

DDBJ Pipeline: Cloud-based analysis tool

researcher





DDBJ supercomputer

no resource? → use DDBJ's supercomputer

no skill? → web-based easy-operation

DDBJ
DNA Data Bank of Japan

http://p.ddbj.nig.ac.jp

DDBJ
DNA Data Bank of Japan

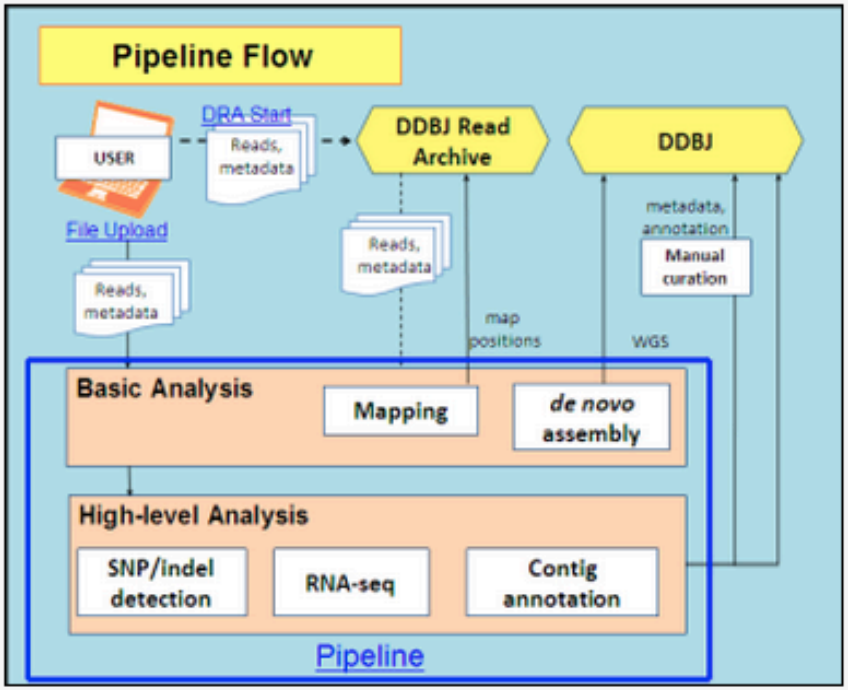
DDBJ Read Annotation Pipeline

English | Japanese

DDBJ Read Annotation Pipeline is a cloud-computing based analytical platform for next-generation sequencing data.

LOGIN

New account | Login as "guest"



User ID:

Password:

Login

Check current jobs

* by the guest account.

Manual & tutorial

- Japanese manual
- English manual
- DBCLS togotv Tutorial video 1 (JP) - Reference Genome Mapping
- DBCLS togotv Tutorial video 2 (JP) - De novo Assembly

Account registration of "DRA"

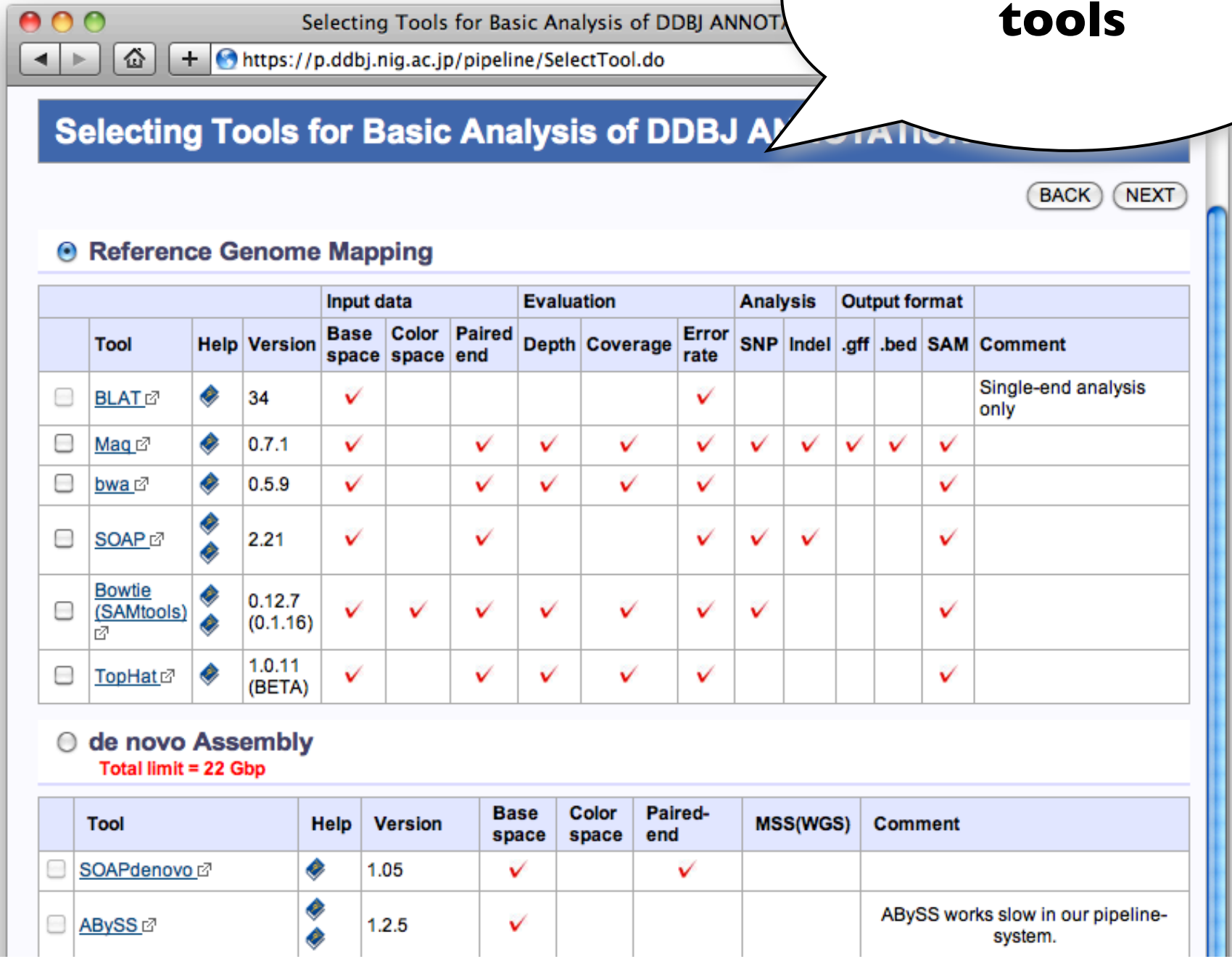
DRA account registration information please see the page.

pipeline
pipeline_info

pipeline_info Reload bugs in 'HTTP upload' function were fixed. Please reload the web page of your uploaded data.

DDBJ pipeline: Software

de *facto* standard tools



Selecting Tools for Basic Analysis of DDBJ ANNOT

https://p.ddbj.nig.ac.jp/pipeline/SelectTool.do

Selecting Tools for Basic Analysis of DDBJ A

BACK NEXT

Reference Genome Mapping

				Input data			Evaluation			Analysis		Output format			
	Tool	Help	Version	Base space	Color space	Paired end	Depth	Coverage	Error rate	SNP	Indel	.gff	.bed	SAM	Comment
☐	BLAT		34	✓					✓						Single-end analysis only
☐	Maq		0.7.1	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	
☐	bwa		0.5.9	✓		✓	✓	✓	✓					✓	
☐	SOAP		2.21	✓		✓			✓	✓	✓			✓	
☐	Bowtie (SAMtools)		0.12.7 (0.1.16)	✓	✓	✓	✓	✓	✓	✓				✓	
☐	TopHat		1.0.11 (BETA)	✓		✓	✓	✓	✓					✓	

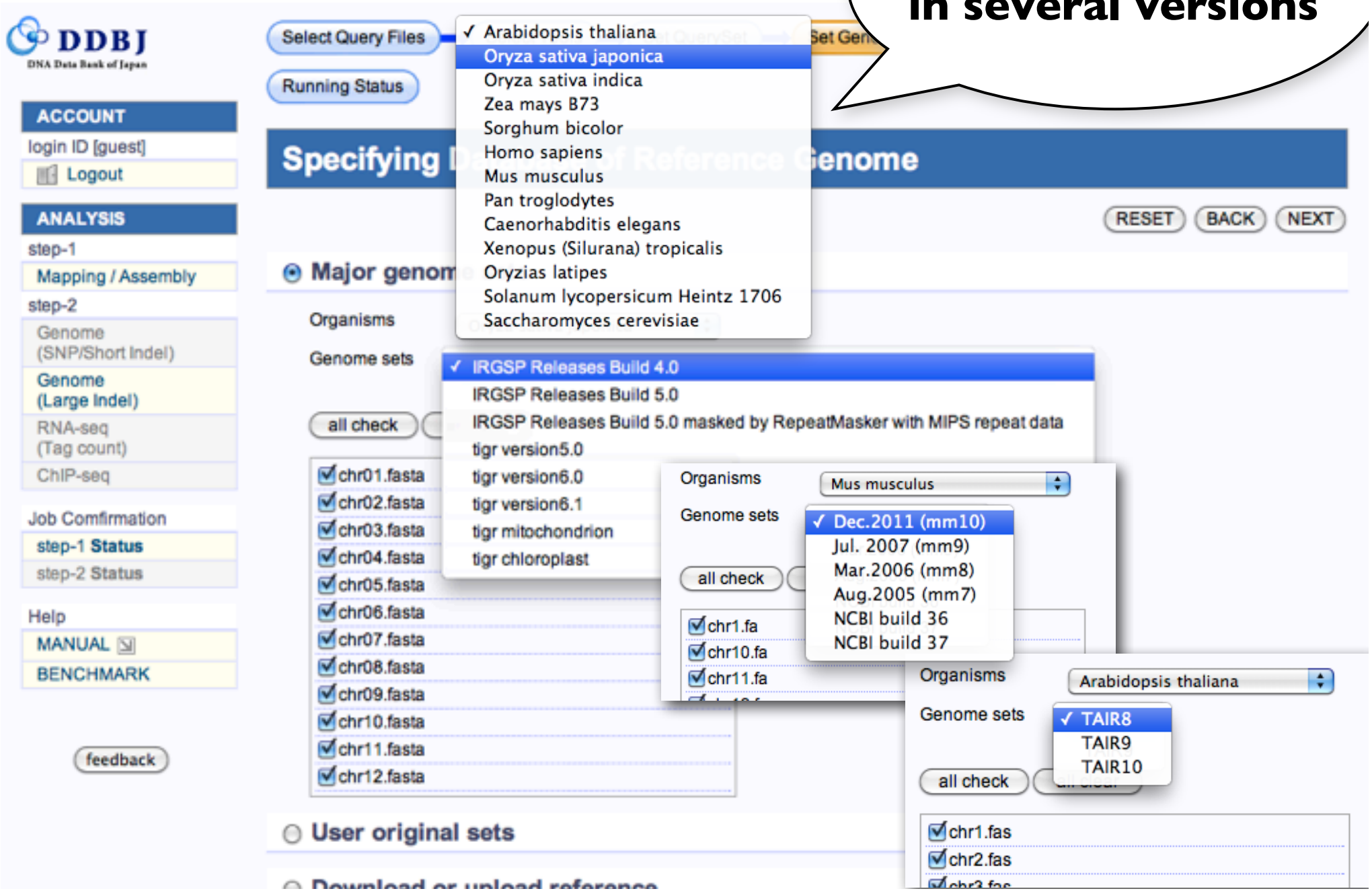
de novo Assembly

Total limit = 22 Gbp

	Tool	Help	Version	Base space	Color space	Paired-end	MSS(WGS)	Comment
☐	SOAPdenovo		1.05	✓		✓		
☐	ABySS		1.2.5	✓				ABySS works slow in our pipeline-system.

DDBJ pipeline: references

major genome sets in several versions



http://www.migap.org

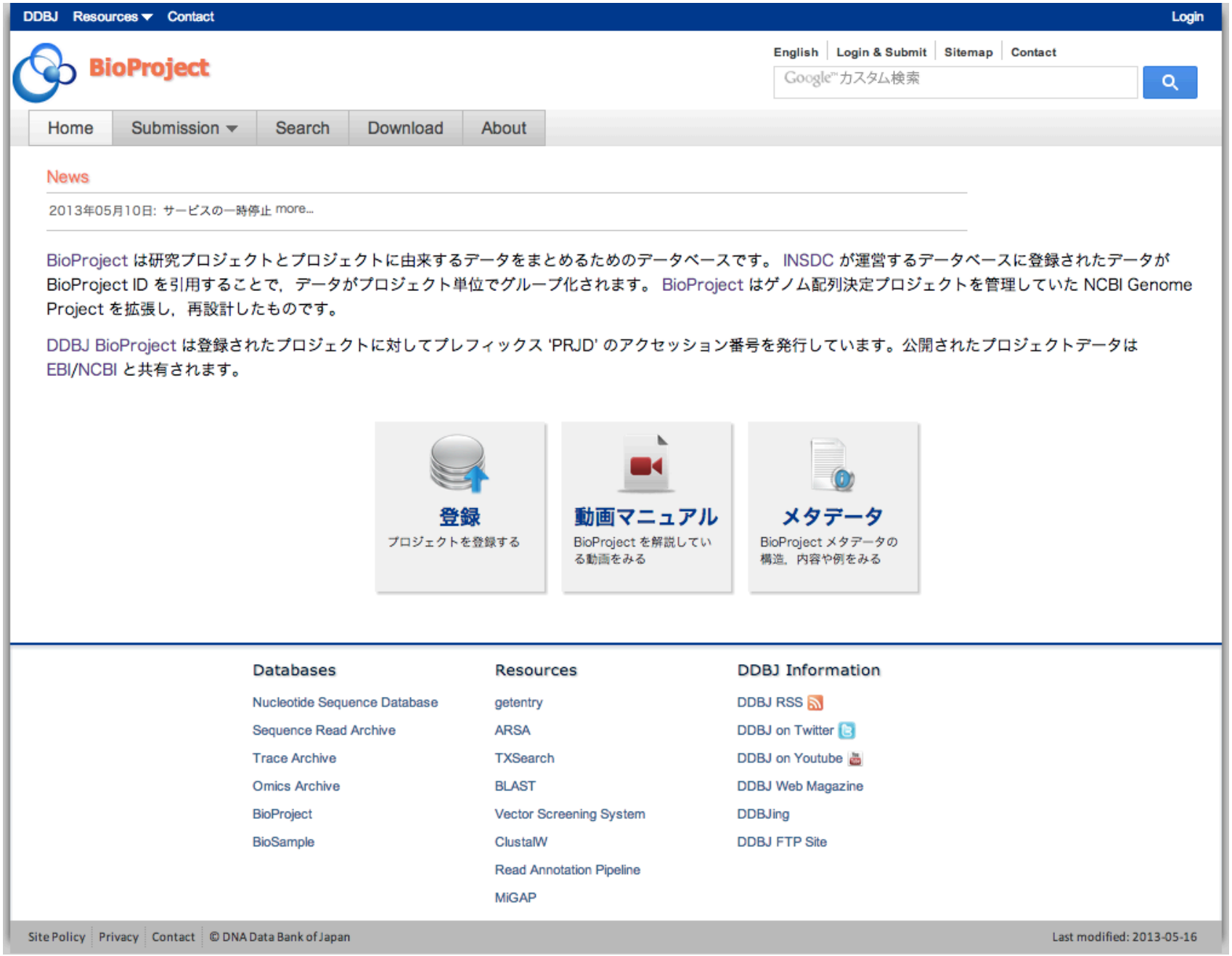
DDBJ
2012-2013

new

BioProject / BioSample DBs in INSDC

- **BioProject** (loanched 2011 in DDBJ)
  - A collection of biological data from a single initiative, an organization or a consortium. The DB provides users a single place to find links to the diverse data types generated for that project.
- **BioSample** (construction in DDBJ)
  - The DB contains descriptions of biological source materials used in experimental assays. The purpose of the database is to provide unified storage and access to information about biological samples. These samples may have information stored in other databases (e.g. nucleotide sequence, expression).

BioProject (started in 2011)



Project Detail : PRJDB36

Accession	PRJDB36
Project Type	Primary submission
Project Data Type	Genome sequencing

General info

Project title	Halomonas sp. KM-1, a highly bioplastic poly(3-hydroxybutyrate)-producing bacterium
Project Description	The Halomonas sp. KM-1 that was isolated in Ikeda city, Japan was deposited in the International Patent Organism Depositary (IPOD, AIST Japan) as FERM BP-10995
Release Date	2012-01-27
Relevance	Industrial

Grant 1

GrantId	K2017, K2161 and K22040
Title	a Grant-in-Aid for scientific research from the Ministry of the Environment of Japan
Agency abbreviation	MOE
Agency	the Ministry of the Environment of Japan

Biomaterial provider

Biomaterial provider	IPOD, FERM BP-10995

Project Type

Sample scope/Material/Capture/Methodology

Sample Scope	Monoisolate
Material	Genome
Capture	Whole
Methodology	Sequencing

Objective

Objective	RawSequenceReads, Analysis

Target

Organism information

Organism name	Halomonas sp. KM-1
Taxonomy ID	590061

Project Publication

Publication 1

PubMed ID	22172913

Title and general info

Grant supplier and material info

Project type and objective

Organism info and publication

refactored

http://www.ddbj.nig.ac.jp/

WordPress
A free content management system (CMS)

DDBJ
DNA Data Bank of Japan

ENGLISH

サイト内検索

HOME | 塩基配列の登録 | 利用の手引き | 検索・解析 | FTP・WebAPI | レポート・統計 | お問い合わせ

- DDBJの紹介
- Q&A集

塩基配列の登録
- SAKURA
- 大量登録システム(MSS)
- データの修正・更新
- DDBJ Sequence Read Archive
- DDBJ Trace Archive

プロジェクトの登録
- DDBJ BioProject Database

検索
- getentry
- ARSA
- TXSearch
- BLAST

系統解析
- ClustalW

DDBJ : DNA Data Bank of Japan

DDBJ（日本DNAデータバンク）は欧州と米国の対応機関(EBIおよびNCBI)と密接に協力しながら DDBJ/EMBL/GenBank国際塩基配列データベースを構築している三大国際DNAデータバンクのひとつです

Hot Topics
一覧へ

- 2012.01.12 DDBJ リリース 88.0, DAD リリース 58.0 完成
- 2012.01.11 ブタ (*Sus scrofa*) full length enriched cDNA 配列 データの公開
- 2012.01.05 Nucleic Acids Research に DDBJ に関する論文発表

Maintenance
一覧へ

- 2011.12.21 （全サービス再開） (2012/1/12-18) DDBJ 公開サービスの一時停止予定

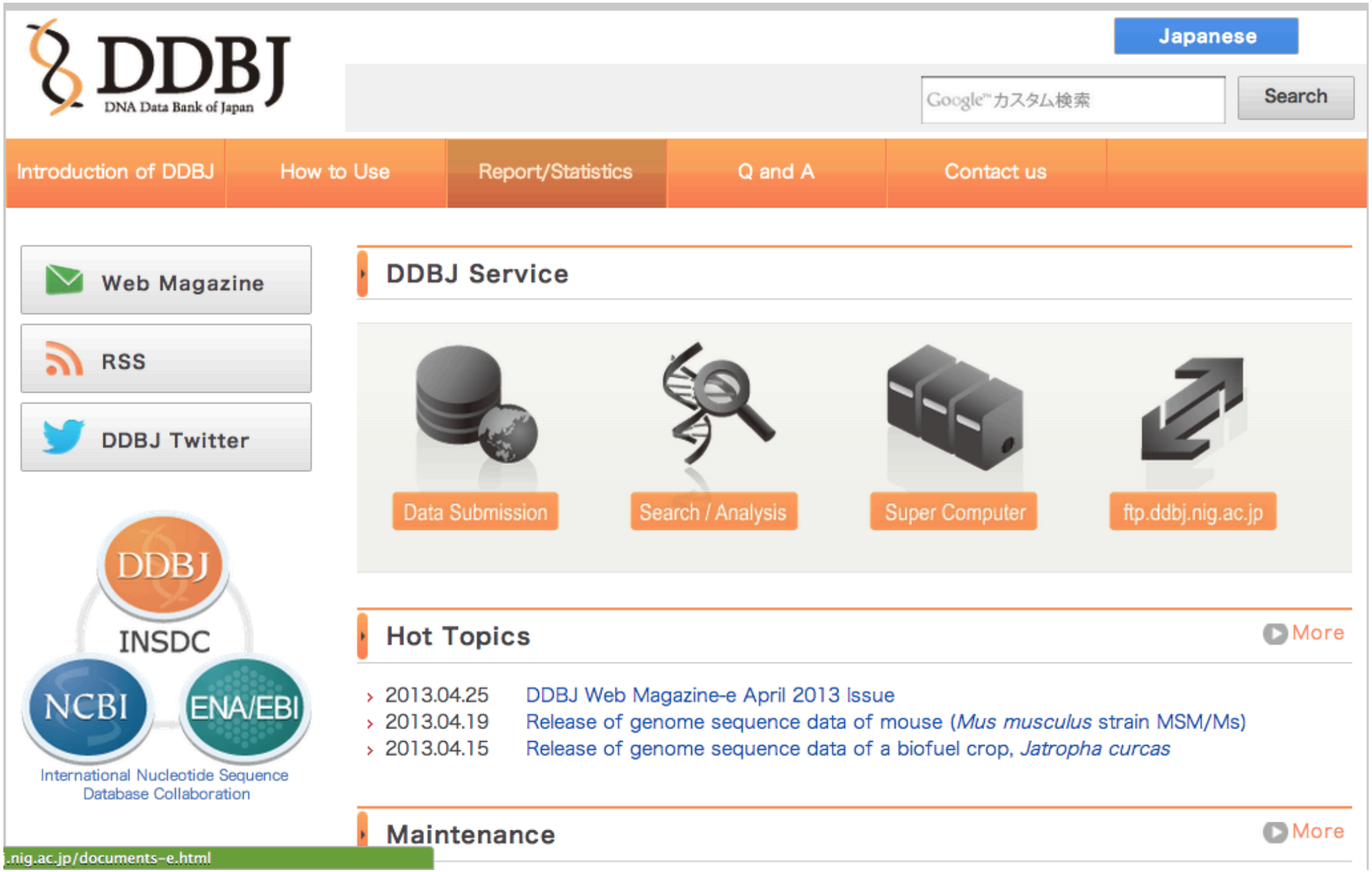
Information

- 「第25回 DDBJing 講習会 in 三島」開催のお知らせ
- DDBJ メールマガジン No.67 配信

塩基配列の登録・更新

FTP・Web API

Renewal of the Web (20 May 2013)



Renewal of submission system (D-easy)

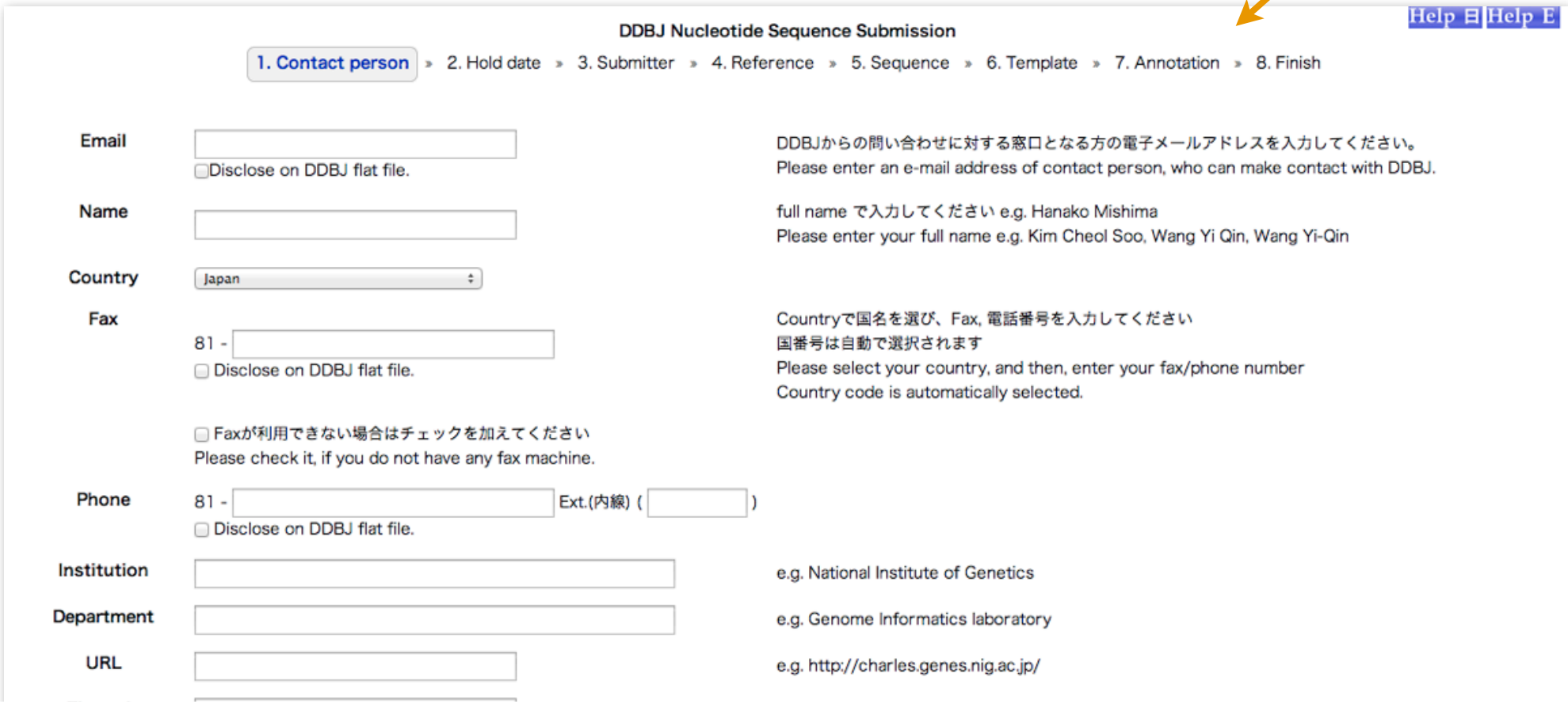
DDBJ submission portal

http://ddbj.nig.ac.jp/submission/

Nucleotide

Submission of small-scale nucleotide sequence data with annotation. In case of project data, please use BioProject, MSS, and DRA.

Create new submission



restored

NIG Supercomputer overview

DDBJ
DNA Data Bank of Japan

Special machines for big memory (*de novo* assembly)

Thin nodes
(16 cores, 64GB mem) ×
352 nodes

Medium nodes
(80 cores, 2TB mem) ×
2 nodes

Fat node
(768 cores, 10TB mem) × 1
node

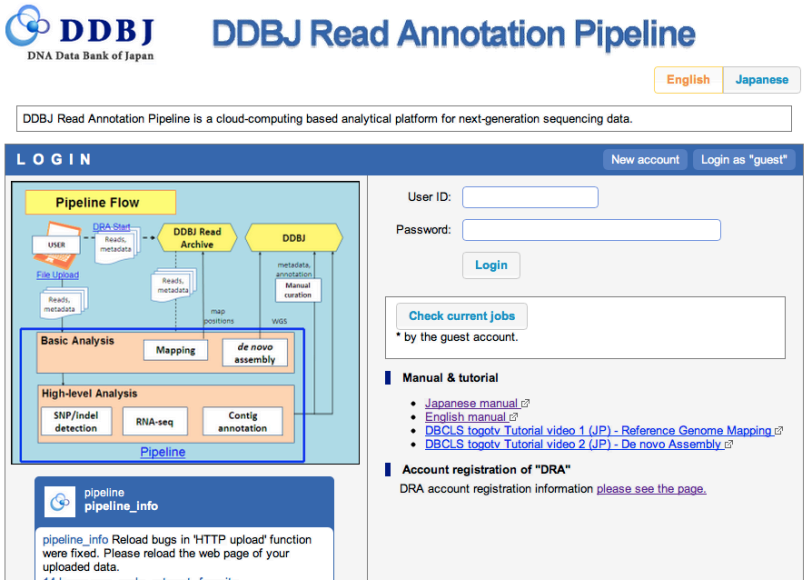
Infiniband QDR
40Gbps

Inter-
connection
network

Fast HDD for
calculation
(Lustre)
2PB

HDD for
archiving
(MAID)
3PB

main DB's getentry/ARSA BLAST/clustalW


DDBJ
DNA Data Bank of Japan
DDBJ Read Annotation Pipeline
English
Japanese
DDBJ Read Annotation Pipeline is a cloud-computing based analytical platform for next-generation sequencing data.
LOGIN
New account
Login as "guest"
Pipeline Flow
DRA Start
USER
Reads, metadata
DDBJ Read Archive
DDBJ
File Upload
Reads, metadata
metadata, annotation
Manual curation
map positions
WGS
Basic Analysis
Mapping
de novo assembly
High-level Analysis
SNP/indel detection
RNA-seq
Contig annotation
Pipeline
pipeline
pipeline_info
pipeline_info Reload bugs in 'HTTP upload' function were fixed. Please reload the web page of your uploaded data.
User ID:
Password:
Login
Check current jobs
* by the guest account.
Manual & tutorial
Japanese manual
English manual
DBCLS togotv Tutorial video 1 (JP) - Reference Genome Mapping
DBCLS togotv Tutorial video 2 (JP) - De novo Assembly
Account registration of "DRA"
DRA account registration information please see the page.


MiGAP
Microbial Genome Annotation Pipeline
HOME
FORUM
FAQ
HELP
INFO
Top Menu
About MiGAP
Help
Information
About Pipeline
MiGAP Server Operation Team
List of articles on MiGAP usage and review
Acknowledgement
Login the Pipeline
Login the Old Pipeline
Home › About MiGAP
What is MiGAP?
Thursday, 31 December 2009 00:00 | Last Updated on Wednesday, 02 May 2012 17:32 | Written by Administrator
The number of bacterial genomes collected by Genome Information Broker of DDBJ [2] has increased every week and will reach 1000 genomes in September 2009. Thanks to the revolution of the sequencing technology, many microbiologists will get genome sequences of their favorite strains. It is certain that we will soon observe tsunami of genome sequences. However, it is uncertain that every microbiologist is able to fully utilize the genome sequence for their research. The biological knowledge will remain only a drop in the ocean, if the bottleneck is not removed. The bottleneck is the annotation. Therefore, we have developed a microbial genome annotation pipeline (MiGAP) to support for novice and old pro alike to interpret sequences.
More than 1,000 microbial complete genomes have been sequenced as of December 2009 and the rate of sequencing will rocket ahead thanks to the 2nd and 3rd generation sequencers. However, the tsunami of sequence data does not necessarily mean the increase of our knowledge on microbes. The sequences have to be annotated. MiGAP (Microbial Genome Annotation Pipeline) provides novice and old pro alike with a mechanical annotation to microbial contigs and genomes.
MiGAP identifies ORFs and RNA regions and infers the functions of ORFs by referring to highly evaluated public databases. MiGAP has the following three modes of the operation:
b-MiGAP provides analysis by the default setting of programs, parameters and the reference databases. The user is required to just give sequences to MiGAP to get the annotation
s-MiGAP provides the user with the freedom to select programs, parameters and the reference databases.
g-MiGAP provides the user with the function of add his/her own tools and databases to the pipeline in addition to s-MiGAP function.
Please get UserID from DNA Data Bank of Japan, National Institute of Genetics to start using MiGAP. The
Recent News
Announcement of MiGAP service suspension due to maintenance
List of articles on MiGAP usage and review
Browsing and downloading of annotation results on the old super computer, are now resumed.
Acknowledgement
Introduction & Practice (as of May 2012)
About malfunction of tRNA prediction in MiGAP
MiGAP service is resumed in the new supercomputer in NIG
Announcement of MiGAP service suspension due to the server replacement
Job throwing is resumed.
Announcement of MiGAP service suspension due to maintenance
search...

new trial

Development of an Ontology for the INSDC Feature Table Definition

DDBJ / NIG

Yaz Nakamura

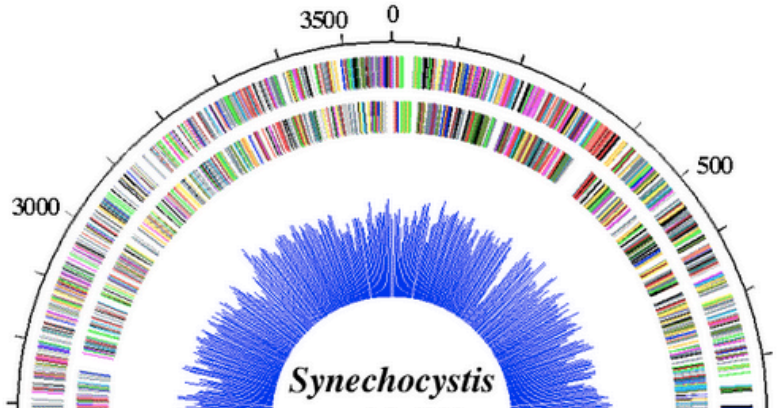
ICM 2013 at EBI

CyanoBase

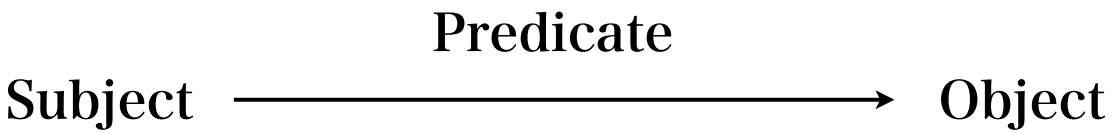
Nakamura, Y., Kaneko, T., Hirosawa, M., Miyajima, N. and Tabata, S. (1998) CyanoBase, a www database containing the complete nucleotide sequence of the genome of *Synechocystis* sp. strain PCC6803. *Nucleic Acids Res.* **26,** 63-67.

Nakamura, Y., Kaneko, T., Miyajima, N. and Tabata, S. (1999) Extension of CyanoBase. CyanoMutants: repository of mutant information on *Synechocystis* sp. strain PCC6803. *Nucleic Acids Res.* **27,** 66-68.

Nakao, M., Okamoto, S., Kohara, M., Fujishiro, T., Fujisawa, T., Sato, S., Tabata, S., Kaneko, T. and Nakamura, Y. (2010) CyanoBase: the cyanobacteria genome database update 2010. *Nucleic Acids Res.,* **38,** D379-381.



RDF (Resource Description Framework)



Gene group ? | +/- | Top

geneset	type	gene_member
Synechocystis/genesets/ppi,sll0520,sll1938	ppi	sll0520, sll1938
Synechocystis/genesets/ppi,sll0520,slr0452	ppi	sll0520, slr0452
Synechocystis_sp._PCC_6803/NADH_dehydrogenase_I	protein_complex	sll0223, sll0520, sll0521, sll0522, sll0026, sll0027, sll0519, slr0851, slr1280, slr2007, slr2009, sll1733, slr0331, slr0844, slr1279, slr1743, sll1484, sll1732, slr0261, slr1281, slr1291, ssr1386

variation ? | +/- | Top

gene id	gene product	gene symbol	variation id	location	variation type	ref_allele	allele
slr0468	hypothetical protein		CynPCC6803-Shestakov_000000014	Chr:2602717	SNP:exonic(slr0468)+downstream(slr0467)	C	A
slr0468	hypothetical protein		CynPCC6803-GT-I_000000011	Chr:2602717	SNP:exonic(slr0468)+downstream(slr0467)	C	A
slr0468	hypothetical protein		CynPCC6803-Shestakov_000000005	Chr:2602734	SNP:exonic(slr0468)+downstream(slr0467)	T	A
slr0468	hypothetical protein		CynPCC6803-PCC-N_000000023	Chr:2602717	SNP:exonic(slr0468)+downstream(slr0467)	C	A
slr0468	hypothetical protein		CynPCC6803-PCC-P_000000007	Chr:2602717	SNP:exonic(slr0468)+downstream(slr0467)	C	A
slr0468	hypothetical protein		CynPCC6803-GT-I_000000012	Chr:2602734	SNP:exonic(slr0468)+downstream(slr0467)	T	A
slr0468	hypothetical protein		CynPCC6803-PCC-N_000000005	Chr:2602734	SNP:exonic(slr0468)+downstream(slr0467)	T	A
slr0468	hypothetical protein		CynPCC6803-PCC-P_000000036	Chr:2602734	SNP:exonic(slr0468)+downstream(slr0467)	T	A

Data integration by RDF technology

DDBJ DNA Data Bank of Japan

Genome DB entries　　curated information　　SNPs from SRA

gene_id	gene_product	gene_symbol	annotation_id	curated_gene_symbol	reference_count	annotation_count	valiation_id	valiation_type	seqid	location	allele
SCO0613	arginine deiminase		http://togo.annotation.jp/annotations/193815	arcA	1	1	SITK24_000002495	SNP	chr	652849	C
SCO0674	endo-1,4-beta-xylanase		http://togo.annotation.jp/annotations/197580	xysA	1	4	SITK24_000002703	SNP	chr	713507	C
SCO0674	endo-1,4-beta-xylanase		http://togo.annotation.jp/annotations/197580	xysA	1	4	SITK24_000002702	SNP	chr	712924	T
SCO0674	endo-1,4-beta-xylanase		http://togo.annotation.jp/annotations/197580	xysA	1	4	SITK24_000002704	SNP	chr	713822	G
SCO0713	lipase		http://togo.annotation.jp/annotations/191409	lipA	1	1	SITK24_000002874	SNP	chr	756500	G
SCO0713	lipase		http://togo.annotation.jp/annotations/191409	lipA	1	1	SITK24_000002875	SNP	chr	756716	A
SCO0713	lipase		http://togo.annotation.jp/annotations/191409	lipA	1	1	SITK24_000002876	SNP	chr	756797	C
SCO0713	lipase		http://togo.annotation.jp/annotations/191409	lipA	1	1	SITK24_000002877	SNP	chr	757109	C
SCO1483	carbamoyl phosphate synthase large subunit	carB	http://togo.annotation.jp/annotations/193838	pyrA	1	5	SITK24_000005845	SNP	chr	1584905	G
SCO1483	carbamoyl phosphate synthase large subunit	carB	http://togo.annotation.jp/annotations/193838	pyrA	1	5	SITK24_000005847	SNP	chr	1586465	C
SCO1483	carbamoyl phosphate synthase large subunit	carB	http://togo.annotation.jp/annotations/193838	pyrA	1	5	SITK24_000005841	SNP	chr	1584152	A
SCO1483	carbamoyl phosphate synthase large subunit	carB	http://togo.annotation.jp/annotations/193838	pyrA	1	5	SITK24_000005842	SNP	chr	1584191	G
SCO1483	carbamoyl phosphate synthase large subunit	carB	http://togo.annotation.jp/annotations/193838	pyrA	1	5	SITK24_000005843	SNP	chr	1584359	A
SCO1483	carbamoyl phosphate synthase large subunit	carB	http://togo.annotation.jp/annotations/193838	pyrA	1	5	SITK24_000005844	SNP	chr	1584890	A
SCO1483	carbamoyl phosphate synthase large subunit	carB	http://togo.annotation.jp/annotations/193838	pyrA	1	5	SITK24_000005846	SNP	chr	1586213	G
SCO1484	carbamoyl phosphate synthase small subunit		http://togo.annotation.jp/annotations/193839	pyrAa	1	2	SITK24_000005848	SNP	chr	1588303	T
SCO1486	dihydroorotase	pyrC	http://togo.annotation.jp/annotations/193846	pyrC	1	1	SITK24_000005849	SNP	chr	1589698	G
SCO1486	dihydroorotase	pyrC	http://togo.annotation.jp/annotations/193846	pyrC	1	1	SITK24_000005850	SNP	chr	1589974	G
SCO1486	dihydroorotase	pyrC	http://togo.annotation.jp/annotations/193846	pyrC	1	1	SITK24_000005851	SNP	chr	1590196	C
SCO1487	aspartate carbamoyltransferase catalytic subunit	pyrB	http://togo.annotation.jp/annotations/193841	pyrB	1	8	SITK24_000005852	SNP	chr	1590420	G
SCO1487	aspartate carbamoyltransferase catalytic subunit	pyrB	http://togo.annotation.jp/annotations/193841	pyrB	1	8	SITK24_000005853	SNP	chr	1590854	A
SCO1513	GTP pyrophosphokinase		http://togo.annotation.jp/annotations/194993	relA	1	3	SITK24_000005907	SNP	chr	1618230	G
SCO1513	GTP pyrophosphokinase		http://togo.annotation.jp/annotations/194993	relA	1	3	SITK24_000005908	SNP	chr	1618512	C
SCO1513	GTP pyrophosphokinase		http://togo.annotation.jp/annotations/194993	relA	1	3	SITK24_000005909	SNP	chr	1618563	G
SCO1513	GTP pyrophosphokinase		http://togo.annotation.jp/annotations/194993	relA	1	3	SITK24_000005906	SNP	chr	1617687	A
SCO1565	glycerophosphoryl diester phosphodiesterase		http://togo.annotation.jp/annotations/194568	glpQ1	1	10	SITK24_000006076	SNP	chr	1675111	T
SCO1565	glycerophosphoryl diester phosphodiesterase		http://togo.annotation.jp/annotations/194568	glpQ1	1	10	SITK24_000006078	SNP	chr	1675304	T
SCO1565	glycerophosphoryl diester phosphodiesterase		http://togo.annotation.jp/annotations/194568	glpQ1	1	10	SITK24_000006079	SNP	chr	1675353	A
SCO1565	glycerophosphoryl diester phosphodiesterase		http://togo.annotation.jp/annotations/194568	glpQ1	1	10	SITK24_000006080	SNP	chr	1675541	G
SCO1565	glycerophosphoryl diester phosphodiesterase		http://togo.annotation.jp/annotations/194568	glpQ1	1	10	SITK24_000006075	SNP	chr	1675042	G
SCO1565	glycerophosphoryl diester phosphodiesterase		http://togo.annotation.jp/annotations/194568	glpQ1	1	10	SITK24_000006077	SNP	chr	1675270	G
SCO1570	argininosuccinate lyase		http://togo.annotation.jp/annotations/193777	argH	1	17	SITK24_000006090	SNP	chr	1681035	C
SCO1570	argininosuccinate lyase		http://togo.annotation.jp/annotations/193777	argH	1	17	SITK24_000006091	SNP	chr	1681419	C
SCO1570	argininosuccinate lyase		http://togo.annotation.jp/annotations/193777	argH	1	17	SITK24_000006087	SNP	chr	1680439	C
SCO1570	argininosuccinate lyase		http://togo.annotation.jp/annotations/193777	argH	1	17	SITK24_000006088	SNP	chr	1680447	T
SCO1570	argininosuccinate lyase		http://togo.annotation.jp/annotations/193777	argH	1	17	SITK24_000006089	SNP	chr	1680944	G
SCO1570	argininosuccinate lyase		http://togo.annotation.jp/annotations/193777	argH	1	17	SITK24_000006092	SNP	chr	1681632	C
SCO1576	arginine repressor		http://togo.annotation.jp/annotations/193728	argR	2	103	SITK24_000006113	SNP	chr	1686916	C
SCO1576	arginine repressor		http://togo.annotation.jp/annotations/193728	argR	2	103	SITK24_000006114	SNP	chr	1687093	T

INSDC seuqence entry file (DDBJ format)

```
LOCUS       AP011615             6788435 bp    DNA     circular HTG 16-APR-2010
DEFINITION  Arthrospira platensis NIES-39 DNA, nearly complete genome.
ACCESSION   AP011615
VERSION     AP011615.1
DBLINK      BioProject:PRJDA42161
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Arthrospira platensis NIES-39
  ORGANISM  Arthrospira platensis NIES-39
            Bacteria; Cyanobacteria; Oscillatoriales; Arthrospira.
REFERENCE   1  (bases 1 to 6788435)
  AUTHORS   Fujisawa,T., Fujita,N. and Sekine,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-NOV-2009) to the DDBJ/EMBL/GenBank databases.
            Contact:Takatomo Fujisawa
            National Institute of Technology and Evaluation, NITE, Bioresource
            Information Center, Department of Biotechnology; 2-49-10
            Nishihara, Shibuya, Tokyo 151-0066, Japan
            URL     :http://www.bio.nite.go.jp/
REFERENCE   2
  AUTHORS   Fujisawa,T., Narikawa,R., Okamoto,S., Ehira,S., Yoshimura,H.,
            Suzuki,I., Masuda,T., Mochimaru,M., Takaichi,S., Awai,K.,
            Sekine,M., Horikawa,H., Yashiro,I., Omata,S., Takarada,H.,
            Katano,Y., Kosugi,H., Tanikawa,S., Ohmori,K., Sato,N., Ikeuchi,M.,
            Fujita,N. and Ohmori,M.
  TITLE     Genomic Structure of an Economically Important Cyanobacterium,
            Arthrospira (Spirulina) platensis NIES-39
  JOURNAL   DNA Res. 17, 85-103 (2010)
COMMENT     Genome Coverage: 11x
            Sequencing Technology: ABI 3730
            The genome structure of A. platensis is estimated to be a single,
            circular chromosome of 6.8 Mb, based on optical mapping.
FEATURES             Location/Qualifiers
     source          1..6788435
                     /db_xref="taxon:696747"
                     /mol_type="genomic DNA"
                     /organism="Arthrospira platensis NIES-39"
                     /strain="NIES-39"
     CDS             152..412
                     /codon_start=1
                     /locus_tag="NIES39_A00010"
                     /product="hypothetical protein"
                     /protein_id="BAI87842.1"
                     /transl_table=11
                     /translation="MFDYSFGFPEAAIAFLGLFSEAAIAFFGLSLEELPGFLKLAFGF
                     FLGFCQFRQQPPNIVSIGGHGGSSQVSQSSILATFHYAIAFW"
     CDS             complement(377..724)
                     /codon_start=1
                     /locus_tag="NIES39_A00020"
                     /product="hypothetical protein"
                     /protein_id="BAI87843.1"
                     /transl_table=11
                     /translation="MLVVMLIDPQNKERSPIASPVRSPSLSEVRSPTSEVRSPTLSEV
                     RSPTFSEVRSRSVPVGRIAYRLASAIAYRLASAIAQRARRAYRLSPRKCIPVASRKCD
                     RILPESDRIVESG"
```

Features/Locations

Feature Annotation Location Description Ontology is developed in BioHackathon 2012, http://biohackathon.org/resource/faldo

Features/Qualifiers

developed in this work

OWL definition for INSDC FT

In Japanese domestic BioHackathon BH12.12, we described using the OWL in INSDC Features/Qualifiers Annotation Description, a part of document in Feature Table Definition (Version 10.2 November 2012)

* defined a feature key as an owl:Classs.

* defined a qualifier key as an owl:DatatypeProperty.

* defined a qualifier value, which is any value from the controlled vocabulary, as range or Individuals of data property.

* defined the relations between Feature-Qualifer as a Class domain of data property

* defined the mandatory and optional qualifier by the restrictions class of feature class.

* the molecular scope and organisim scope of Feature as obeject property.

OWL: Web Ontology Language

A trial - https://github.com/tfuji/INSDC

DDBJ DNA Data Bank of Japan

INSDC / insdc.ttl

tfuji a month ago Create insdc.ttl

1 contributor

file | 15629 lines (14443 sloc) | 569.432 kb — Edit | Raw | Blame | History

```
1  @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
2  @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
3  @prefix owl: <http://www.w3.org/2002/07/owl#> .
4  @prefix : <http://insdc.org/owl/#> .
5  @prefix xml: <http://www.w3.org/XML/1998/namespace> .
6  @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
7
8  <http://insdc.org/owl/>
9      a owl:Ontology ;
10     rdfs:label "insdc" ;
11     rdfs:seeAlso "http://www.insdc.org/" ;
12     owl:equivalentClass [
13         owl:maxCardinality "1"^^xsd:nonNegativeInteger ;
14         owl:onProperty <http://insdc.org/owl/allele>
15     ], [
16         owl:maxCardinality "1"^^xsd:nonNegativeInteger ;
17         owl:onProperty <http://insdc.org/owl/citation>
18     ], [
19         owl:maxCardinality "1"^^xsd:nonNegativeInteger ;
20         owl:onProperty <http://insdc.org/owl/db_xref>
21     ], [
22         owl:maxCardinality "1"^^xsd:nonNegativeInteger ;
23         owl:onProperty <http://insdc.org/owl/experiment>
24     ], [
25         owl:maxCardinality "1"^^xsd:nonNegativeInteger ;
26         owl:onProperty <http://insdc.org/owl/gene>
```

Querying in SPARQL (Query)

DDBJ
DNA Data Bank of Japan

Virtuoso SPARQL Query Editor

About | Namespace Prefixes | Inference rules

Default Data Set Name (Graph IRI)

http://insdc.org/

Query Text

```
select ?feature_label, ?qualifier_label, ?qualifier_type,?qualifier_comment
where {
?s owl:annotatedSource ?feature;
   owl:annotatedTarget [
        a owl:Restriction ;
        owl:onProperty ?qualifier
    ] .
?s rdfs:isDefinedBy ?o.
?qualifier rdfs:label ?qualifier_label.
?o rdfs:label ?qualifier_type.
?qualifier rdfs:comment ?qualifier_comment.
?qualifier rdfs:domain ?feature.
?feature rdfs:label ?feature_label.
FILTER(?feature_label = "source")
} order by ?qualifier_type limit 15
```

Show qualifiers for the source feature (limit 15)

Querying in SPARQL (Answer)

DDBJ DNA Data Bank of Japan

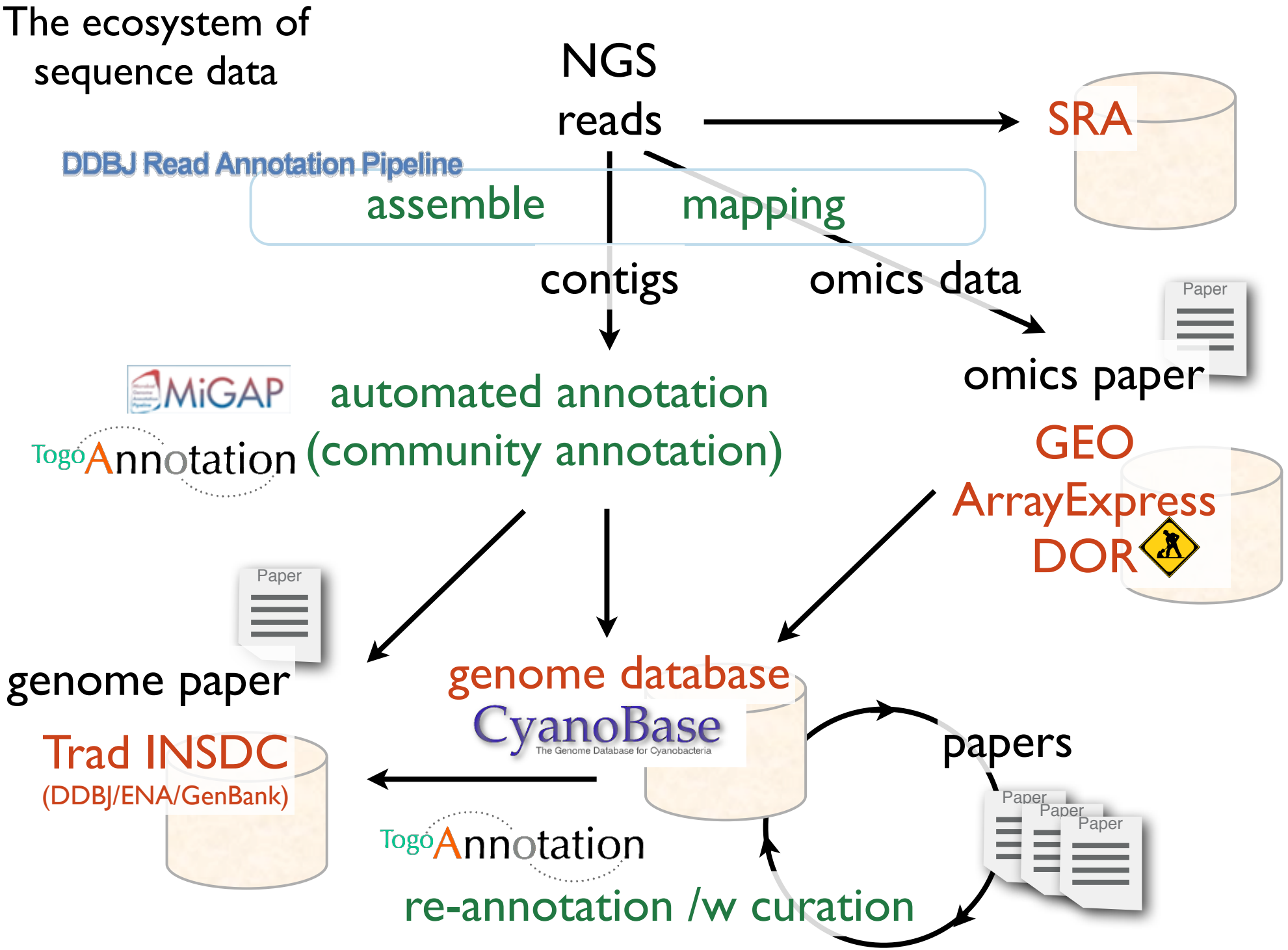
2 mandatory qualifiers and 47 optional qualifiers

feature_label	qualifier_label	qualifier_type	qualifier_comment
source	mol_type	Mandatory qualifiers	in vivo molecule type of sequence
source	organism	Mandatory qualifiers	scientific name of the organism that provided the sequenced genetic material.
source	chromosome	Optional qualifiers	chromosome (e.g. Chromosome number) from which the sequence was obtained
source	plasmid	Optional qualifiers	name of naturally occurring plasmid from which the sequence was obtained, where plasmid is defined as an independently replicating genetic unit that cannot be described by /chromosome or /segment
source	ecotype	Optional qualifiers	a population within a given species displaying genetically based, phenotypic traits that reflect adaptation to a local habitat.
source	focus	Optional qualifiers	identifies the source feature of primary biological interest for records that have multiple source features originating from different organisms and that are not transgenic.
source	germline	Optional qualifiers	the sequence presented in the entry has not undergone somatic rearrangement as part of an adaptive immune response; it is the unrearranged sequence that was inherited from the parental germline
source	haplogroup	Optional qualifiers	name for a group of similar haplotypes that share some sequence variation. Haplogroups are often used to track migration of population groups.
source	isolate	Optional qualifiers	individual isolate from which the sequence was obtained
source	lab_host	Optional qualifiers	scientific name of the laboratory host used to propagate the source organism from which the sequenced molecule was obtained
source	lat_lon	Optional qualifiers	geographical coordinates of the location where the specimen was collected
source	macronuclear	Optional qualifiers	if the sequence shown is DNA and from an organism which undergoes chromosomal differentiation between macronuclear and micronuclear stages, this qualifier is used to denote that the sequence is from macronuclear DNA.
source	mating_type	Optional qualifiers	mating type of the organism from which the sequence was obtained; mating type is used for prokaryotes, and for eukaryotes that undergo meiosis without sexually dimorphic gametes
source	mol_type	Optional qualifiers	in vivo molecule type of sequence
source	organelle	Optional qualifiers	type of membrane-bound intracellular structure from which the sequence was obtained

This ontology provides a set of classes, properties, and restrictions that can be used to represent and interchange biological feature information generated in different systems and under different contexts.

Future plan

- We plan to extend the application this ontology:
  - The data integration of the genome annotation and relative resources and new web application using semantic web technologies.
  - Configuration and validation on new DDBJ Nucleotide Sequence Submission System.


The ecosystem of sequence data
NGS reads
SRA
DDBJ Read Annotation Pipeline
assemble
mapping
contigs
omics data
Paper
omics paper
MiGAP
TogoAnnotation
automated annotation
(community annotation)
GEO
ArrayExpress
DOR
Paper
genome paper
genome database
CyanoBase
The Genome Database for Cyanobacteria
papers
Paper
Paper
Paper
Trad INSDC
(DDBJ/ENA/GenBank)
TogoAnnotation
re-annotation /w curation