Jun Mashima, Takehide Kosuge, Toshihisa Okido, Yuichi Kodama, Kazuho Ikeo, Hideaki Sugawara, Yoshio Tateno and Takashi Gojobori Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, Japan



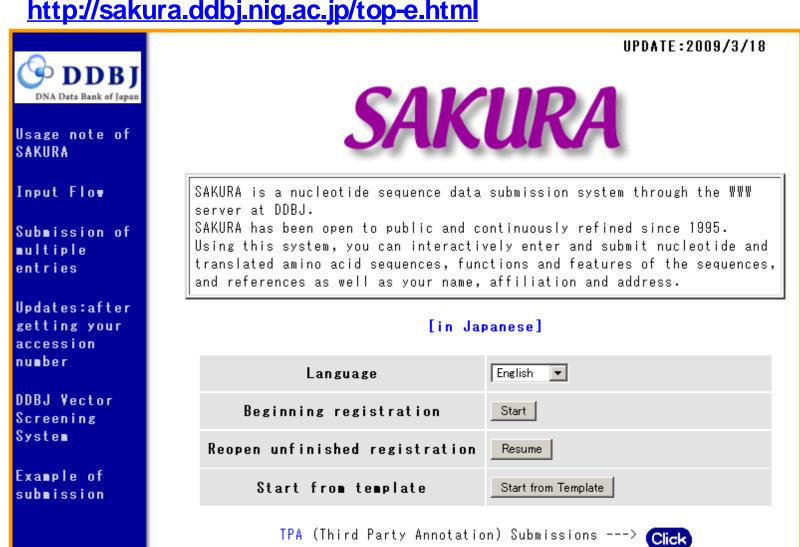
DNA Data Bank of Japan

Nucleotide Sequence Data INSDC (DDBJ/EMBL-Bank/GenBank)

SAKURA

Interactive application for nucleotide sequence submission to enter all of items required for submission on a step-by-step basis

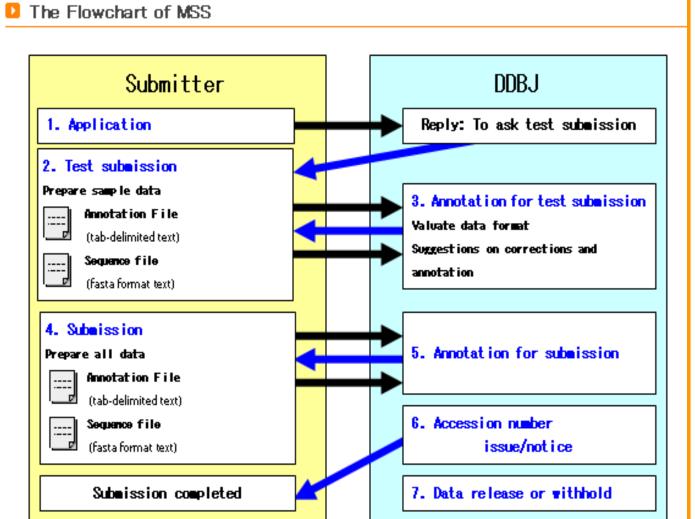
http://sakura.ddbj.nig.ac.jp/top-e.html



MSS (Mass Submission System)

- Large number of sequences (e.g. EST data)
- Complex submission with many features
- Long sequences (e.g. genome)
- Data for WGS, MGA etc.

http://www.ddbj.nig.ac.jp/sub/mss_flow-e.html



DDBJ is a member of INSDC. While many biological databases are constructed on data derived from the primary literature, most of sequence data submitted to INSDC are directly submitted from general researchers as a prerequisite to publication (required by most journals).

In direct submissions from general biologists there are many errors not only formative or careless mistakes but also scientific or logical misunderstandings. So, to find them manual curation is one of the most important jobs for DDBJ. In principle, DDBJ annotators review all sequence data submitted to DDBJ to suggest error corrections for submitters.

.

History of Annotation Jamborees DDBJ Annotators Participated/Coorganized

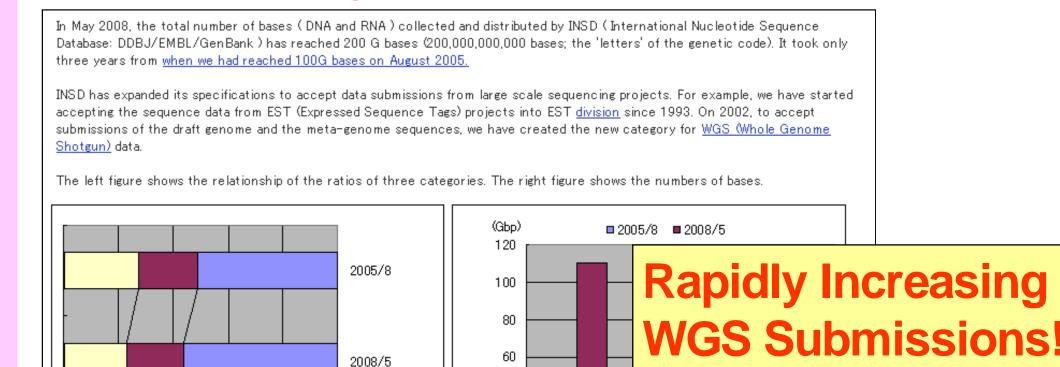
FANTOM: Mouse cDNA	2000	DD
FANTOM2: Mouse cDNA	2002	coll
H-Inv1: Human cDNA	2002	anr
E. coli Bioinformatics Meeting	2003	
H-Inv2: Human cDNA	2003	H-I
E. coli annotation workshop	2003	we
RAP1: Rice genome and cDNA	2004	DD
E. coli annotation workshop II	2005	
RAP2: Rice genome and cDNA	2006	
<i>Theileria</i> genome annotation	2007-2008	

DBJ annotators sometimes ollaborate with large scale nnotation projects. Invitational and H-Invitational 2 ere coorganized by JBIRC and

Nucleotide Number Collected by INSDC

 $1986 \sim 2005 \, 100 \, \mathrm{Gbp}$ over $1986 \sim 2008 \, 200 \, \mathrm{Gbp}$ over

x 2, only in 3 years!



http://www.ddbj.nig.ac.jp/whatsnew/2008/080606-e.html

■ WGS ■ EST/GSS

Gene Expression Data

CIBEX

Center for Information Biology gene EXpression database

- Public repository for microarray data - Curated resource for data browsing and download

http://cibex.nig.ac.jp

GIB

Comprehensive

genomes in the

public domain

data repository of

complete microbial

GTPS

Reannotated

GIB data by

categorized

grade

Genome Information Broker

PDDB

GIB-M

GTPS

Gene Trek in Prokaryote Space

S was updated on Mar 25, 2009.

Aeropyrum pernix KI

owse genomes (List of microbe genomes)

Тор	Search	Experiment	List A	ray Lis	t :	Submission	Contact
	AME-compliant	ase for microarra : data in accorda		ED Soci		CIE	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Search		Brow	rse			Link	
Experim Accessior Fitle Authors	_	arrays	summary list of in CIBEX. riment List / List	all experir	ments or	MGED socie MGED hon MIAME MAGE	-
advanced se	arch	GO Statis	stics riments	:	31	Microarray I ArrayExpre	
Array		Array	'S	:	51	Other Datab	ases
Accessior Model Nar		Hybri	dizations	: odated 200	1077	DDBJ DNA DNA RATIE OF JA	and a
Provider			idol up	wateu zut	7070070	Submission	
opics	earch	GO				If you want to sub microarray data, p submission site o bex	olease visit our
		the submission s		ata form	ing tool.	Grant Suppor	t
Note:	If the warning dia accept our site c					The construction is supported by "(Publication of Sci Results" from JSF	of this database Grant-in-Aid for entific Research

Submission to DDBJ

http://gib.genes.nig.ac.jp

Search selected genomes at once

chaea; Crenarchaeota

obaculum arsenaticum DSM 13514

Sulfolobus acidocaldarius DSM 639

Archaea; Euryarchaeota

Halobacterium salinarum R

loquadratum walsbyi DSM 16790

anococcoides burtonii DSM 6242

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

page which explains the procedure of analysis was also opened to the public

Internet Explorer 5.x, 6.x 7.x JavaScript and cookies are required to

Gene Trek in Prokaryote Space

S is acronym of Gene Trek in Prokaryote Space. GTPS is a database which is re-annotated against

Genome Information Broker

Hyperthermus butylicus DSM 5456

Pyrobaculum calidifontis JCM 11

Halobacterium sp. NRC-)

Halorubrum lacusprofundi ATCC 49239 Methanocaldococcus jannaschii DSM 266

Methanococcus aeolicus Nankai-3

obaculum aerophilum str. /#

Raw Read Data

Trace/Short Read Archives Trace Archive (TA) and

Short Read Archive (SRA)



🎐 DDB1

About DDBJ

■ Sequence Submission

Mass Submission

Data Updates

▶ <u>SRS</u>

alv

9

Structure and Function

TXSearch

▶ PSI-BLAST

▶ <u>H-Inv_DB</u>

▶ <u>PMD</u>

→ Links

▶ Protein Analysis

Phylogenetics

Data Bank of Japas

Submission

DDBJ top page http://www.ddbj.nig.ac.jp/

DDBJ: DNA Data Bank of Japan

DDBJ (DNA Data Bank of Japan) is one of three summit

national Nucleotide Sequence Database, through close

27,216 entries

DDBJ Rel. 77 Completed

DAD Rel. 46.0 Released

<u>SRS(Sequence Retrieval S</u>

databanks that construct DDBJ/EMBL/GenBank Inter

collaboration with EBI in Europe and NCBI in USA

Hot Topics

Mar 24, 2009

Mar 19, 2009.

Mar 1, 2009

Maintenance

Mar. 19, 2009

Feb. 26, 2009

Dec. 26, 2008

Seguence Data Submission

<u>Submit my sequences</u>

Updatt my entries

Orientation for the data submission

Guidance for the update of the entry

DNA Data Bank of Japan (DDBJ)

DDBJ exchanges data via the SINET3 computer network.

DDBJ banner images

FTP/WebAPI

"Next Inflation" of **Data Submissions Would Be Short Reads** Outputted from 454, Solexa, SOLiD, etc.

Accession DNA Protein AIDBs Taxonomy Site Search

Release of new tomato full length cDNA (HTC) data 10,959 entries and EST data

<u>) service terminated</u>

FTP/ Web API

FTP (ftp.ddbj.ni.

Programmatic interfaces of DDBJ Web services

► Center for Information Biology and DNA Data Bank of Japan (CIB-DDBJ) □

National Institute of Genetics (NIG)

Research Organization of Information and

▶ Department of Genetics □

Download data files

SOKENDAI

₩eb API

Release of new human CAGE MGA 7,824,360 entries NEW

-Invitational DB temporary down

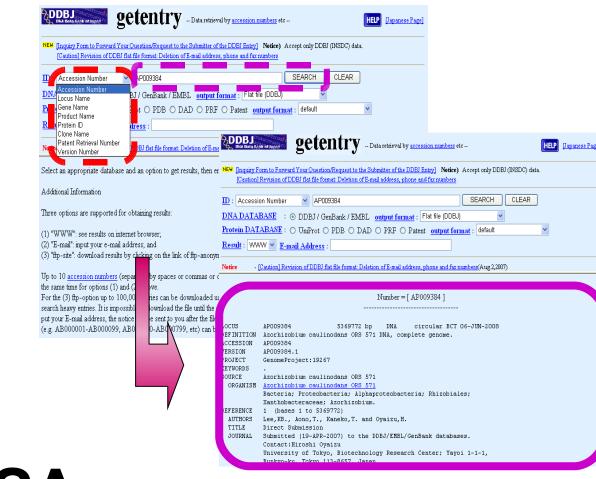
ODBJ C UniProt C PDB C DAD C PRF C Patent >>more

port/Statistics | Contact Us

Japanese

getentry

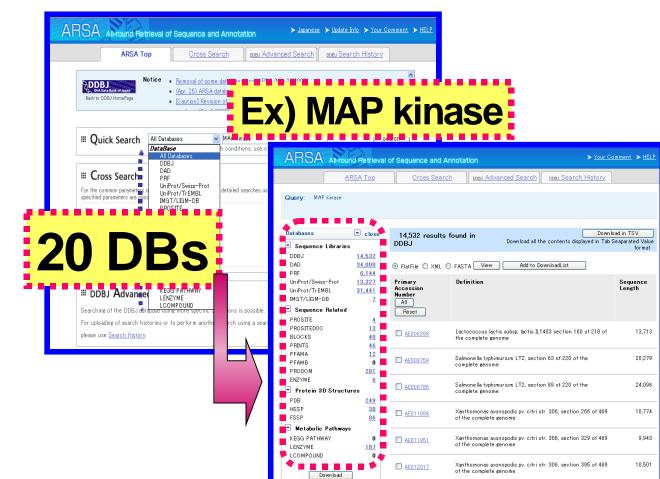
Data retrieval by accession numbers, etc. http://getentry.ddbj.nig.ac.jp/top-e.html



ARSA

All-round Retrieval of Sequence and Annotation

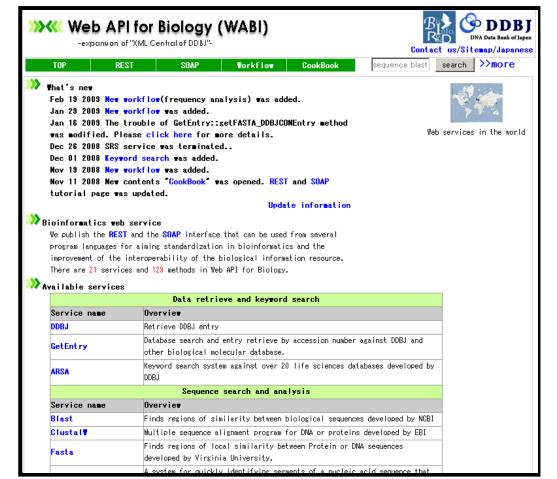
http://arsa.ddbj.nig.ac.jp/top-e.html



WABI (Web API for Biology)

Web Application Program Interface

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



Homology Search

FASTA | BLAST | PSI-BLAST | SSEARCH http://www.ddbj.nig.ac.jp/search/top-e.html

HMMPFAM

A motif search program against Pfam on the basis of Hidden Markov Model (HMM)

http://hmmpfam.ddbj.nig.ac.jp/top-e.html

H-InvDB (mirroring) Database of full-length cDNAs assigned functional annotation as a result of **H-Invitational**

Netscape 6.x, 7.x for Windows

Netscape 6.x, 7.x for Mac OS

http://hinvdb.ddbj.nig.ac.jp/ahg-db/index.jsp

GTOP

Genome to protein structure and function http://spock.genes.nig.ac.jp/~genome/gtop.html

PMD

Site Map

Copyright@DN

Protein Mutant Database http://pmd.ddbj.nig.ac.jp/~pmd/pmd.html

Data Bank of Japan. All Rights Reserved.

Phylogenetics ClustalW

ne Network Platform 📮

t modified: Mar. 30, 2009

Multiple alignment http://clustalw.ddbj.nig.ac.jp/top-e.html

TXSearch

Retrieval of unified taxonomy database

http://txsearch.ddbj.nig.ac.jp/top-e.html