



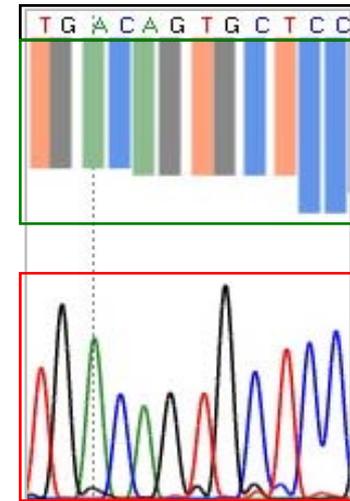
# 塩基配列アーカイブの データベース構築と統合への貢献

大学共同利用機関法人  
情報・システム研究機構  
五條堀 孝



# DDBJ Trace Archive

1. キャピラリ式シーケンサからの出力データ  
(配列、品質情報、波形データ)のためのアーカイブ



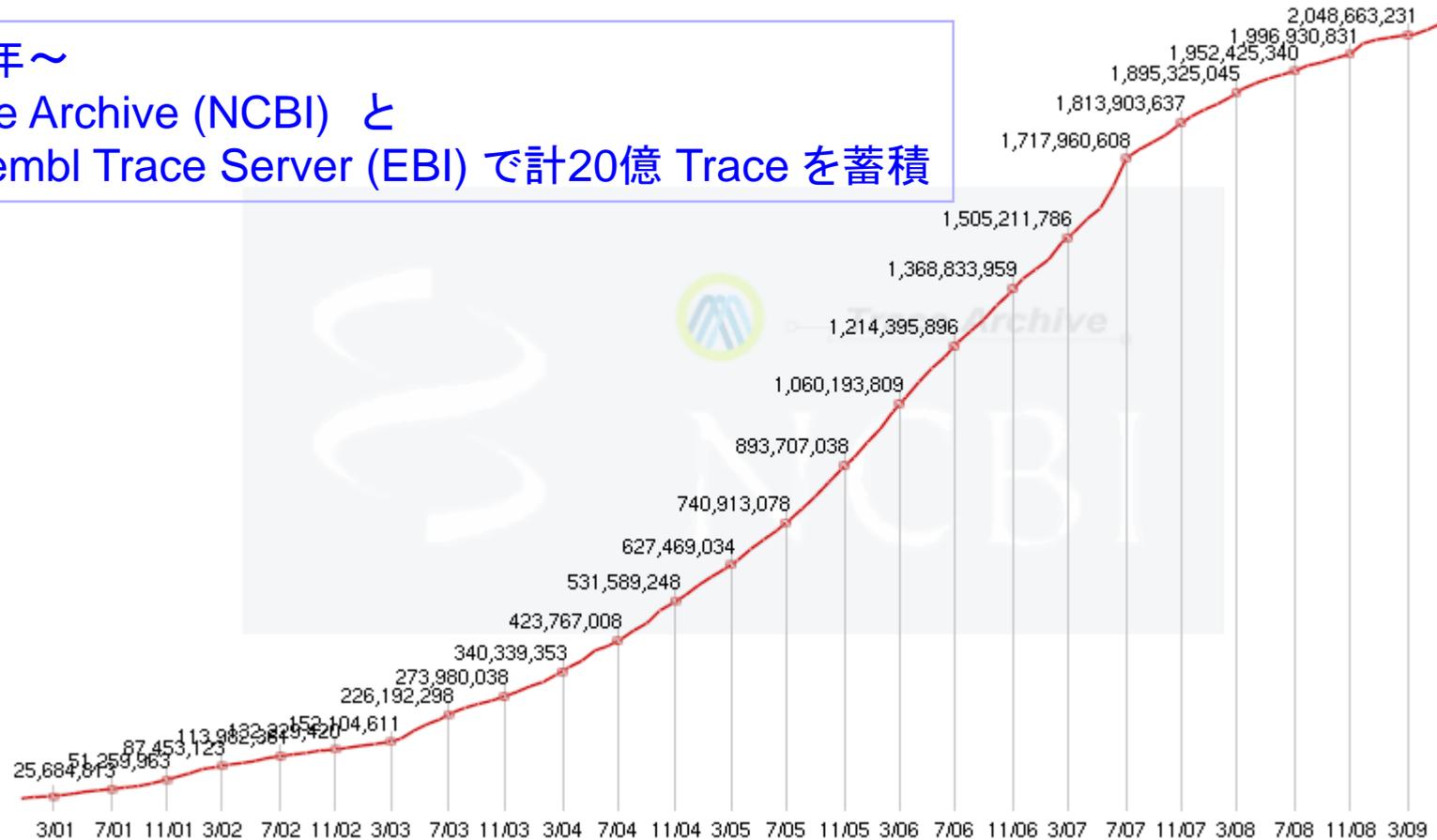
2. 「XML形式のメタデータ」と「データ そのもの」は分離されている
3. 登録データに対して TI(Trace Identifier) 番号を発行  
➡ プレフィックスのないTI番号はNCBIで一括管理されている

# Trace Archive 登録件数

As 2009/05/19 of NCBI Trace Archive contains 2,082,928,978 traces

2001年～

Trace Archive (NCBI) と  
Ensembl Trace Server (EBI) で計20億 Trace を蓄積



➡ 次世代シーケンサへのシフトが進み、登録数の伸びは鈍化

## メタデータ

```

ti: 304753779
trace_name: A01_f_GC1W_70_001.ab1
center_name: NIG
species_code: CIONA INTESTINALIS
strategy: WGA
trace_type_code: MGS
submission_type: NEW
source_type: GENOMIC
taxid: 7718
center_project: GC1W
chemistry_type: TERMINATOR
clip_quality_left: 37
clip_quality_right: 568
clip_vector_left: 37
clip_vector_right: 568
vector_accession: L08752
vector_code: PUC18
plate_id: GC1W_70
primer: GTAAAAAGAGCCAGT
program_id: PHRED
run_date: Sep 8 2001 12:00AM
run_machine_type: 3700
template_id: A01_GC1W_70_001
trace_end: FORWARD
trace_format: ABI
well_id: A01
replaced_by: 0
basecall_length: 907
load_date: Nov 11 2003 9:27AM
state: active
    
```

## 配列

## 品質情報

## 波形データ

NCBI Site map All databases PubMed Search

### Trace Archive

Main Obtaining Data Statistics Tracking Documentation Trace Assembly SRA Trace Home Trace BLAST

Search Searching Tips Searchable Fields Registered Species Submitting Centers FTP

Enter a **query string** (hide Query Builder) or **T1 number**

Submit CENTER\_NAME = "NIG" and SPECIES\_CODE = "HAEMOPHILUS SOMNUS 129PT"

Add and SPECIES\_CODE = HAEMOPHILUS SOMNUS 129PT

Search result: found 395,470,466 items

Your request is: CENTER\_NAME = "WUGSC"

Save result of search as trace .tar .gz file. Only 40,000 traces will be saved (see Searching Tips if you wish to save)

Save query's result All FASTA Quality SCF Mate Pair

Retrieve

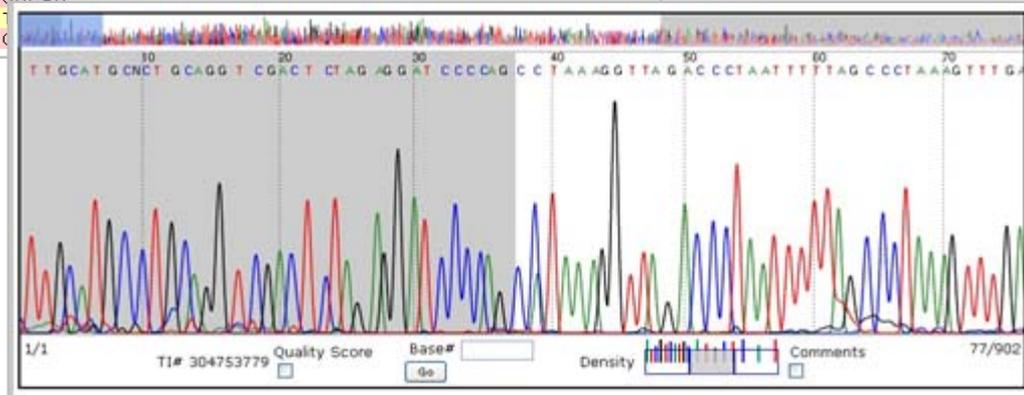
Show as FASTA in color by 5 item(s) per Page 1 / 79,094,096 >>

>gnl|ti|FASTA Quality Info available =>0 - <20 >=20 - <40 >=40 - <60 >=60 - <80 >=80 - <100

name: [jea17d09.b1](#) Mate pair: [5739313](#) [Send to BLAST](#)

```

CCCCCCCC...GTTTTGGTGCCCTTTATTAGCTGCTCGNACGGCCGCCAGTGTGATGGATTCCCAG
ACTGAGAGATTAGTCTCACAAAGCTCAGCTAGATAGCTCTAACTTTGCATTTCAAGGAAGGCAGCAGAGG
CTCAAAAAGAACCTGCCGGCTCCCATGTGCTGCTCCTTGCATTGGCCTTGCCTTCCCTGAGAGGTGAGAG
AGCCACTGTGGTTTGGCTGGCCCTCCCCCTTGGCTATTGGAGTTAGCCTTAAAGCCAGTTATTCTCTGG
AGATTTACAGCCTAGATTTCACTTTTGTGTTGTCTAGATCACGTCTAGTTAGATATTGCCTCTAAGAAA
AGTGGAGGGATATGCTGCAGCTAAAATAGAAGTCTGAAAACCACTGTTCCGGTGTGTCGCTCAGCACTC
ACAGATCTCAAGCTGTATGCACGTAGGACTGGCTGCTGCTGTTTCAGCCTACTGACCCGGAGCTCAGCATAG
AACCTGGTCTGGAAGGAGCTCGGGTGTGCATTTCTGTGGTGTCTGATCTAACTTTCCATAGCCAGGCAG
AGGTGAGGCTCGGACCAGACAGTCTCTTAACCTAACCTGAGACTCACCGATCTGGCTACCTCCCA
GAATCTAAGGCTGGGGTGTCCCAGTGCCTGAATGTCATTTTCTTGTATTNTCCAAAGACTCACGC
    
```



➡ 各種データの閲覧、検索、ダウンロードが可能

## 2008年～ DDBJ Trace Archive

項番	機関名	生物種	機種	件数	容量 (圧縮時)	TI発行日	作業期間	状況
1	遺伝研 小原研	メダカ	ABI	約148万 件	約50GB	7月上旬	約4ヶ月	登録、公開完了
2	東大新領域 服部研	ヒトメタゲノム	ABI	約106万 件	約40GB	7月上旬	約3ヶ月半	登録、公開完了
3	非公開	非公開	ABI	約167万件	不明	未完了		対応中

➡ 国内の研究者に登録を呼びかけていく

## DDBJ Trace Archive

### Search Condition

TI :

CENTER\_NAME :

SPECIES\_CODE :

STRATEGY :

TRACE\_TYPE\_CODE :

<= LOAD\_DATE <

Format :  View Count :

### [Advanced Search](#)

### Registered Data

No.	Organism	Center Name	TI
1	Oryzias latipes	NIG	2095022956-2095389675, 2095396176-2096435759, 2096858496-2096933759
2	Human Metagenome	UTC0B	2097946941-2099007079

<http://trace.ddbj.nig.ac.jp/Trace/search>

➡ DDBJ受付データに対して、検索サービスを提供

## メタデータ

```
ti: 2095022956
trace_name: A01_f_GOLW_10
center_name: NIG
species_code: ORYZIAS LATIPES
strategy: WGA
trace_type_code: WGS
submission_type: NEW
source_type: GENOMIC
taxid: 8090
center_project: MEDAKA_WGS
chemistry: BIGDYE3.0
chemistry_type: TERMINATOR
insert_flank_left: TTGTAACACGACGGCCAGTGCCAAAGCTTGCATGCCTGCAGGTGCGACTCTAGAGGATCCCC
insert_flank_right: GGGTACCGAGCTCGAATTCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTTGTTATC
insert_size: 2250
insert_stdev: 900
plate_id: GOLW_10
primer: GTTTTCCAGTCACGACGTT
primer_code: M13F
program_id: ABI-3.6
run_date: Jul 19 2002 12:00AM
run_machine_id: N027
run_machine_type: ABI3700
seq_lib_id: GOLW_LOT1
strain: HD-RR
svector_accession: L08752
svector_code: PUC18
template_id: A01_GOLW_10
trace_end: FORWARD
trace_format: SCF
well_id: A01
replaced_by: 0
basecall_length: 902
load_date: Jul 2 2008 11:38AM
state: active
```

## 配列 品質情報

## 波形データ

### DDBJ Trace Archive

[Back to Query](#)

Search result: found **1481568** items

TI :  
CENTER\_NAME : NIG  
SPECIES\_CODE :  
STRATEGY :  
TRACE\_TYPE\_CODE :  
LOAD\_DATE :

[Save query's result](#) (Max 40,000 items)

FASTA  Quality  Info  SCF

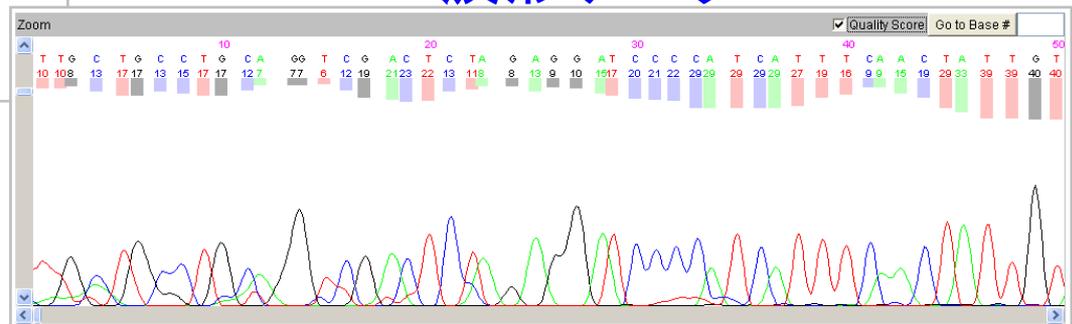
[View](#)

Format : **FASTA**

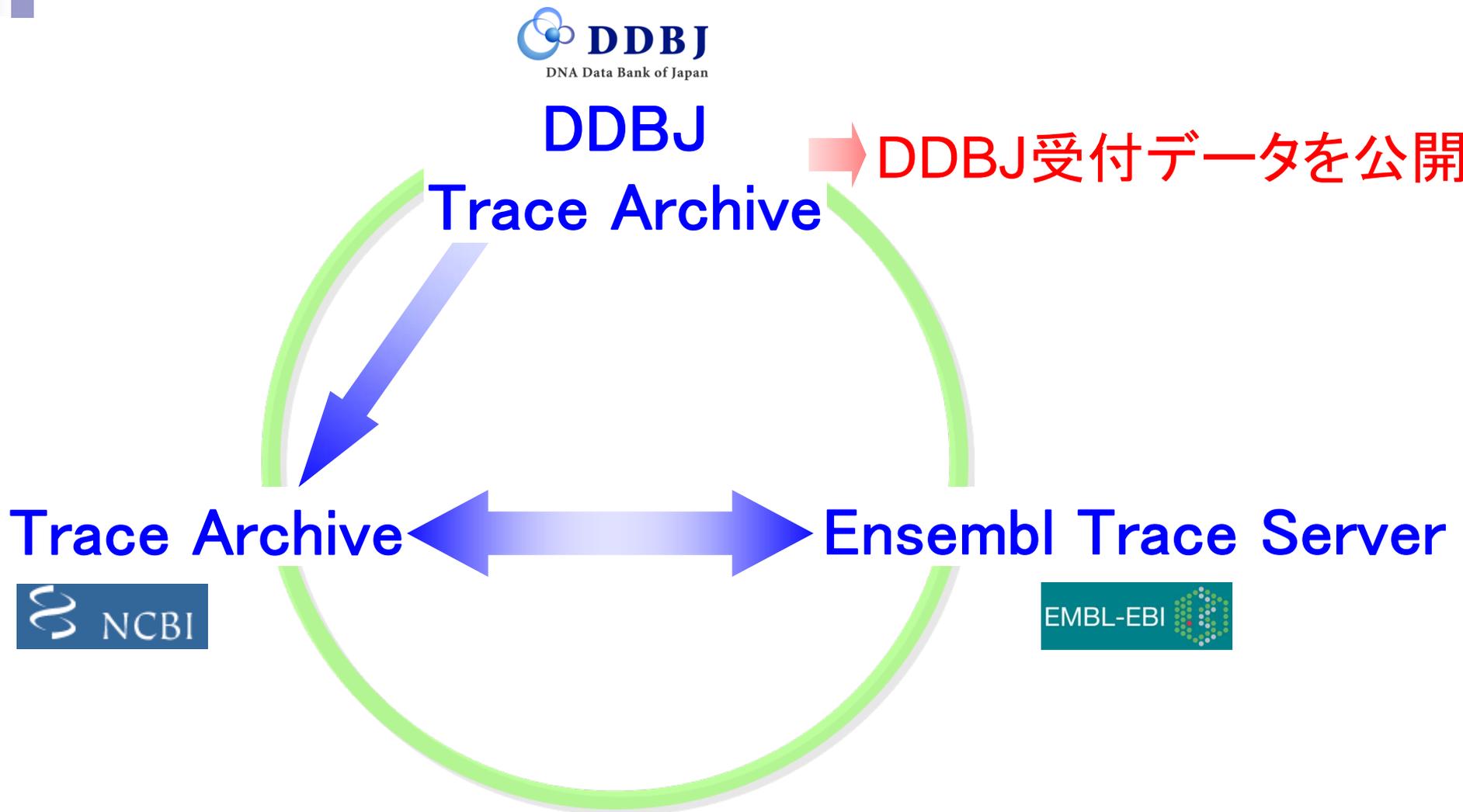
View Count : **5** / Page : **1** / 296314 >>

[Trace](#)

```
>gn|ti|2095022956 name:A01_f_GOLW_10 mate_pair:2095023006
TTGCTGCCTGCAGGTCGACTCTAGAGGATCCCATCATTCAACTATTGTAAGGCGCTCTTAAATTAT
GATTCTGTCGTTTTAGCCAAAAGCAAAAAAAGTATCGTTTTGTTAGAACATAGTTTTCTAGGGAGCGGC
AGTAATCAGAGCAATCCAGCTGTACAGCTTACAGTCCAGATGCCAGACGAGGAAAACGATGACATCCCTGGA
TGTAGTCAGATGCAAAATAGATGCAAACTCTGCGAATTTTCACTTTATAAATGCGTTTTTCTTTTTTAT
CTGCTTCTGATTCACAACAATTTGAAAAAGAAAATACTCAGAAATGCACATTTTTCTTTCTATATGTCTT
CCAACATGAAACACCAAAAAACAATTTTCACTCGGAGTGGGTCTTTAATAAAAAACTCAATTTTTAAT
ATTTTCAAGGCTCATATGAAACACAAAAATAGGTTTTAGCACCAAACTCTTCTTTTAACTGATAACC
AAATTCAGTTTTCAACCATCACAGCTGCTTCTCACTGCAGCTTCAGCTCCATATCTCTACTCTAAGTCA
GGTTTTCTGACAAATGATGGGTCTCCATGCTCTCAATCACTGTGTTTCTGACAAAAAGGTTAAAAAAAT
AGATTCACACTGACCTGGATGTTGACAAAAGAGCCATGGTAGGTTCTCATACAGAACTTTGTTGCCTTTG
TGAGCAAAGGGAACCTCAAAGGGGATCTCAGTTTTACCTCCGGGGACTTTCCAGCCTTGGCAACCTCGAT
GTTGCTGCTGATCANCTGGATGGGCTGCNAGACATTACAGCATAGTACTACGAGCACTGGAGCCAATAT
TTATAATGTGGTCTTNTAGAAAAAAAATACAGTCTGATTAACCGCAAACCGGAAAGGGTGG
```



➔ DDBJ受付データの、各種情報の閲覧、ダウンロードが可能



NCBI と EBI は Trace Archive のさらなる開発は行わない模様  
DDBJ は登録データを増やすことで、アーカイブ構築に貢献する

# Short Read Archive

1. 次世代シーケンサ (454, Solexa, SOLiD etc) の  
“生”出力データ (Instrumentation Data) のためのアーカイブ  
→ 本当の生データである画像データ (Image Data) はサイズ  
が巨大すぎ、アーカイブしていない
2. 加工処理 (アダプター配列の除去 etc) していないデータを登録
3. 「XML形式のメタデータ」と「データそのもの」は分離されている
4. 登録データに対してアクセッション番号を発行

## 次世代シーケンサからの “出力データ (Instrumentation Data)”

454

Data Series	Description
.sff	Flowgram (base call, phred quality score, flow value)

### Solexa

_int.txt	uncalibrated signal intensity values
_nse.txt	uncalibrated noise profiles
_seq.txt	Base calls per read
_prb.txt	Per channel log odds quality scores
_sig2.txt	Phase-corrected signal intensity values

もしくは、

fastq ファイル (Sequence + Qualities)

配列データのみは登録不可

### fastq

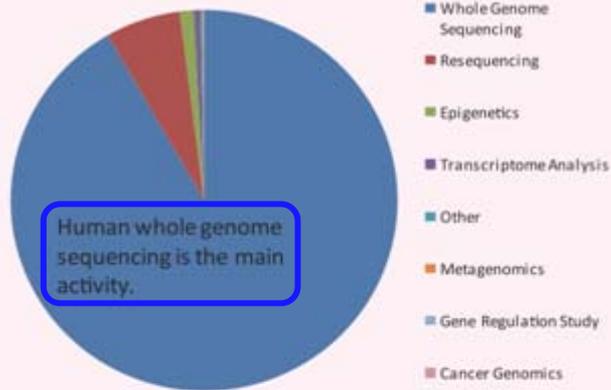
```
@SRR001654.1 9460:7:1:830:763 length=36
GTCAATATTAATCATACCAATATACTCAAAAAATAA
+SRR001654.1 9460:7:1:830:763 length=36
|+~&*4)%+5'#%/)&$%$#%"#&%'"$%#%#!""
@SRR001654.2 9460:7:1:402:781 length=36
GGTCTAAAAAGCAAATTTCAGTCTTCAAATAATTC
+SRR001654.2 9460:7:1:402:781 length=36
||+(%$+% '&+*-0+/*("%&+""&)(*$""#%&$$
```



**10ヶ月程度で5兆塩基に達する見込み  
データサイズ ~100 TB (5/1)**

## Total Bases by Sequencing Study Type

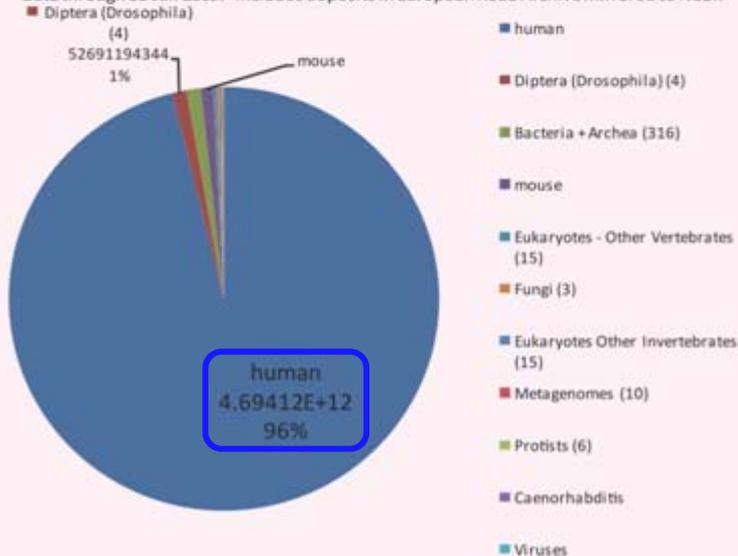
Data through 31 Jan 2009. Includes deposits in European Read Archive mirrored to NCBI.



➔ Whole Genome Sequencing  
が大部分

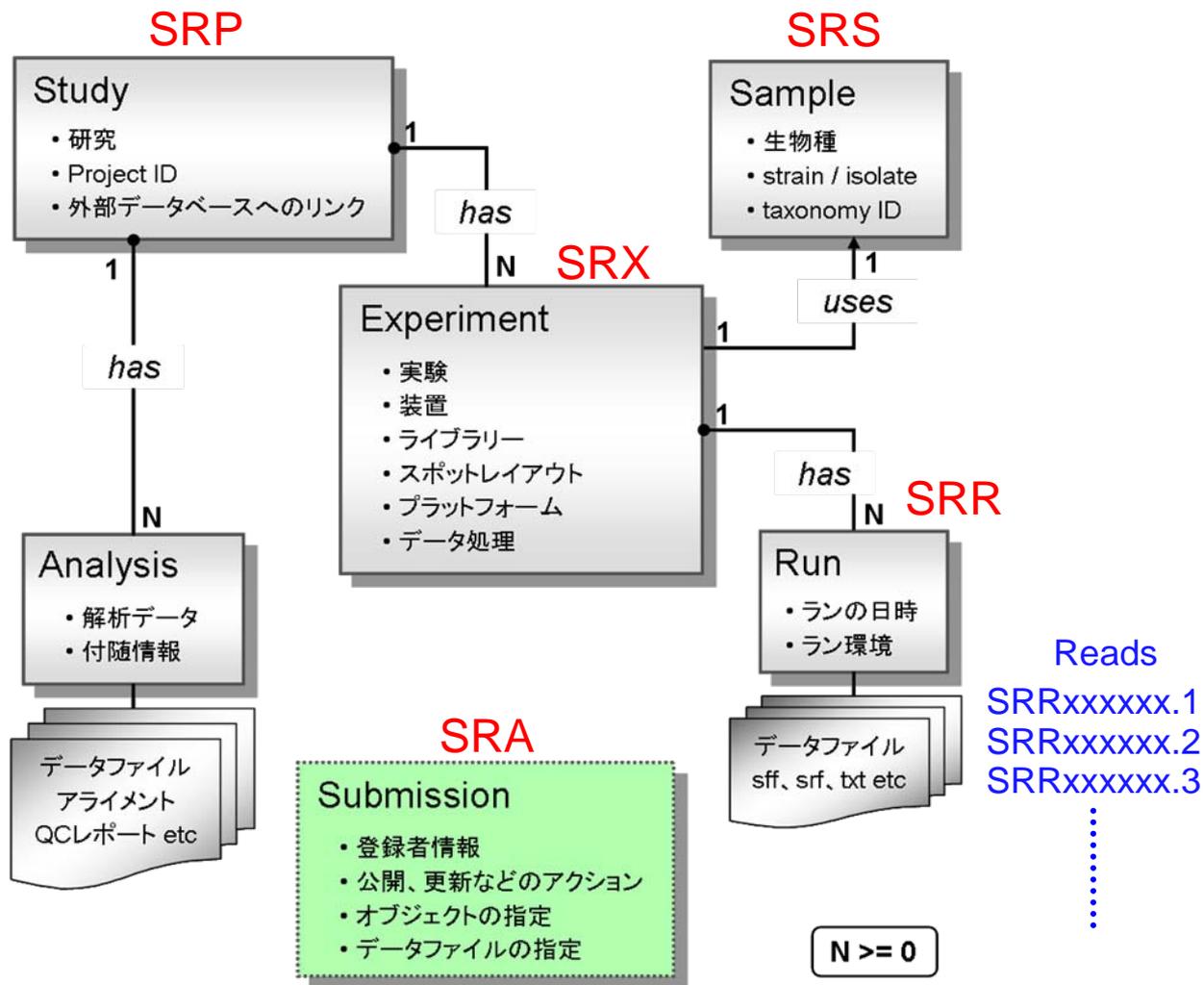
## Bases (species) by Taxonomic Division

Data through 31 Jan 2009. Includes deposits in European Read Archive mirrored to NCBI.



➔ Human が大部分  
「1000人ゲノム」の影響

Metadata (XML) + Data



Submission, Study, Experiment, Sample, Run の  
各オブジェクトにアクセッション番号が発行される

# 公開システム1 (SRA)

## Short Read Archive

Main Browse Search Download Submit Documentation Software Trace Archive Trace Assembly Trace Home Trace BLAST

Study Sample Run Browser Entrez SRA Experiments Entrez Pubmed Entrez GEO DataSets Entrez Genome Project Entrez Taxonomy

### List of Studies

#	Accession	ProjectID	Study Name	Center	SGRP	Center Project
1	<a href="#">ERP000001</a>		Saccharomyces Genome Resequencing Project	SC	SGRP	
2	<a href="#">ERP000002</a>	<a href="#">30943</a>	Complete Genome of a Clinical Isolate of Salmonella enterica serovar Paratyphi A	SC		Salmonella enterica serovar Paratyphi A strain AKU
3	<a href="#">ERP000003</a>		Sequencing flow-sorted X chromosomes from a human female sample (CEU:NA07340) using the Illumina Genome Analyzer	ILLUMINA		X chromosome sequencing project
4	<a href="#">SRP000001</a>	<a href="#">33627</a>	Paired-end mapping reveals extensive structural variation in the human genome	454MSC		HUMAN_SV
5	<a href="#">SRP000002</a>	<a href="#">19655</a>	Allstipes putredinis DSM 17216 Whole Genome Sequencing Project	WUGSC		1970209096
6	<a href="#">SRP000003</a>	<a href="#">19659</a>	Anaerotruncus colihominis DSM 17241 Whole Genome Sequencing Project	WUGSC		1975478781
7	<a href="#">SRP000004</a>	<a href="#">19859</a>	Bacteroides stercoris ATCC 43183 Whole Genome Sequencing Project	WUGSC		1986985037

8 [SRP000005](#)

Short Read Archive

browse Search Download Submit Documentation Software Trace Archive Trace Assembly Trace Home Trace BLAST

Study Sample Run Browser Entrez SRA Experiments Entrez Pubmed Entrez GEO DataSets Entrez Genome Project Entrez Taxonomy

### SRP000005 Biomphalaria glabrata Transcriptome Study

Study Type: Transcriptome Analysis  
Submission: [SRA000138](#) by WUGSC on 2007-06-20T00:00:01Z  
Abstract: none provided  
Description: none provided  
Properties:  
INSDC Project id: [12879](#)

[Download fastq for entire study](#)

#### Experiments

[Show RUNs for each experiment](#)

Accession	Spots	Bases
Total: 3	704,022	184.6M
<a href="#">SRX000011</a>	378,572	99.3M
<a href="#">SRX001379</a>	88,278	21.8M
<a href="#">SRA000138</a>	237,172	63.5M

Study のメタデータ

属している Experiment

## Run Browser

Experiment: **SRX000011**  
454 Sequencing of Biomphalaria glabrata EST Library

属しているRunのメタデータ

Run:

Accession: **SRR000072**   
Alias: ERBRDQF01  
Instrument model: 454 GS FLX  
Date of run: 2007-06-26T15:16:28Z  
Run center: WUGSC  
Statistics:  
Number of spots: 164772  
Number of reads: 329544

Other:

Study: Biomphalaria glabrata Transcriptome Study  
Design: 454 Sequencing of Biomphalaria glabrata EST Library  
Platform: LS454  
Sample: Biomphalaria glabrata  
Library Name: 2010920094  
Library Strategy: FL-cDNA  
Library Source: NON GENOMIC  
Library Selection: cDNA  
Library Layout: SINGLE  
Library Construction Protocol: none provided

Find spots:  X:  Y:    View:  reads (customize)  signals  intensity graph

## Runに含まれるリードの情報

1. **SRR000072.1**

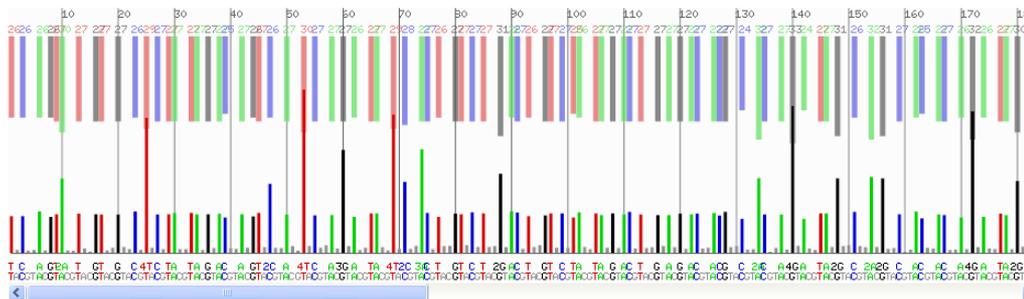
name: ERBRDQF01EGP9U  
plate:ERBRDQF, region:1, x:1713, y:2672  
2. **SRR000072.2**  
name: ERBRDQF01ERSX5  
plate:ERBRDQF, region:1, x:1839, y:3259  
3. **SRR000072.3**  
name: ERBRDQF01D7OXC  
plate:ERBRDQF, region:1, x:1610, y:2910  
4. **SRR000072.4**  
name: ERBRDQF01DVO9H  
plate:ERBRDQF, region:1, x:1474, y:531  
5. **SRR000072.5**  
name: ERBRDQF01EV4MF  
plate:ERBRDQF, region:1, x:1889, y:213  
6. **SRR000072.6**  
name: ERBRDQF01E0UJW  
plate:ERBRDQF, region:1, x:1806, y:538  
7. **SRR000072.7**  
name: ERBRDQF01DVPDA  
plate:ERBRDQF, region:1, x:1470, y:1932  
8. **SRR000072.8**  
name: ERBRDQF01D7KJ  
plate:ERBRDQF, region:1, x:1609, y:65

### Reads (joined)

>gn|SRA|SRR000072.1 ERBRDQF01EGP9U  
tcagTAAATGTGCTTTTCTATAGACAGTCCATTTTCAGGGATATTTTCCAAACTGTCTGGGA  
CTGTCTATAGACTGAGACACGCACAGGGGATAGGCCAAGGCACACAGGGGATAGGN

← 配列

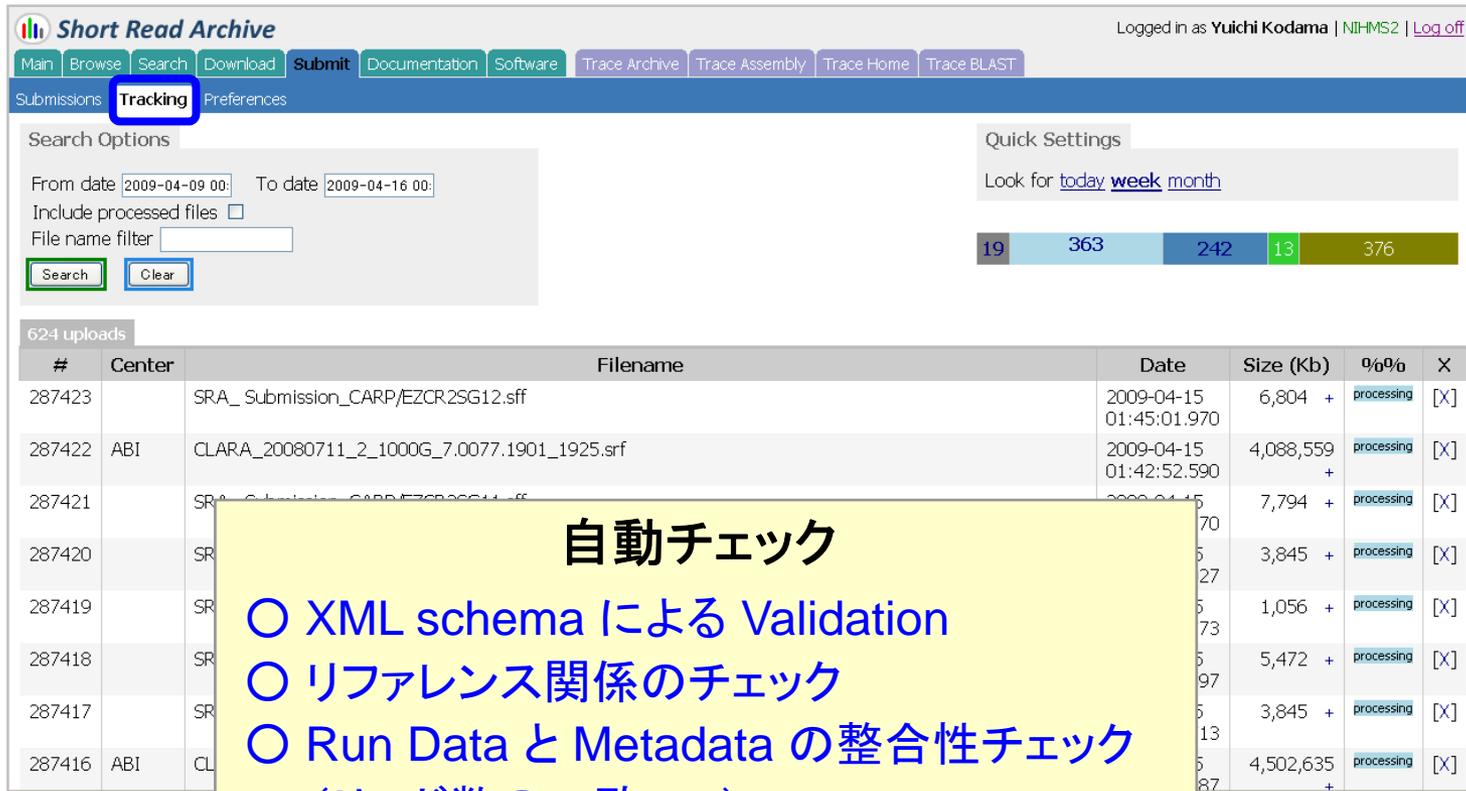
### Intensity graph



← Quality Score

← Intensity

# 登録システム1 (SRA)



Short Read Archive

Logged in as Yuichi Kodama | NIHMS2 | [Log off](#)

Main Browse Search Download Submit Documentation Software Trace Archive Trace Assembly Trace Home Trace BLAST

Submissions **Tracking** Preferences

Search Options

From date: 2009-04-09 00: To date: 2009-04-16 00:  
Include processed files:   
File name filter:

Quick Settings

Look for: [today](#) [week](#) [month](#)

19 363 242 13 376

624 uploads

#	Center	Filename	Date	Size (Kb)	%%	X
287423		SRA_Submission_CARP/EZCR2SG12.sff	2009-04-15 01:45:01.970	6,804	+ processing	[X]
287422	ABI	CLARA_20080711_2_1000G_7.0077.1901_1925.srf	2009-04-15 01:42:52.590	4,088,559	+ processing	[X]
287421		SRA_Submission_CARP/EZCR2SG11.sff	2009-04-15 01:45:01.970	7,794	+ processing	[X]
287420	SR		2009-04-15 01:45:01.970	3,845	+ processing	[X]
287419	SR		2009-04-15 01:45:01.970	1,056	+ processing	[X]
287418	SR		2009-04-15 01:45:01.970	5,472	+ processing	[X]
287417	SR		2009-04-15 01:45:01.970	3,845	+ processing	[X]
287416	ABI	CL	2009-04-15 01:45:01.970	4,502,635	+ processing	[X]

## 自動チェック

- XML schema による Validation
- リファレンス関係のチェック
- Run Data と Metadata の整合性チェック (リード数の一致 etc)

用意されたディレクトリに Aspera (高速ファイル転送ソフト) でアップロード



“Metadata + Run Data”

↓ 自動チェックを通過したデータは Submission 画面に現れる

Short Read Archive

Logged in as Yuichi Kodama | NIHMS2 | [Log off](#)

Main Browse Search Download **Submit** Documentation Software Trace Archive Trace Assembly Trace Home Trace BLAST

**Submissions** Tracking Preferences

SRA submission list for Yuichi Kodama

Create new submission

### Curator による査定

(意味的、自動化できないチェック etc)

Attention (17) Complete (0)

Accession #	Submission #	Submitter	Updated	State	Status	Comments
<a href="#">SRA008162.3</a>	DDBJ : UT-MGS_29	✔ DDBJ	2009-03-13 10:41	error	15	<ul style="list-style-type: none"><li>• <a href="#">SRR013727.1</a> :: data-load: run file problems failed, error code 35584, message:/sra/staging/loading/SRR013727/run.sh: line 3: 21172 Segmentation fault /sra/bin/sradb -q -rc /sra/.sradbrc -run SRR013727 -platform ILLUMINA -rps 1 -batch batch.cfg -log run.log</li></ul>
<a href="#">SRA008167.3</a>	DDBJ : UT-MGS_34	✔ DDBJ	2009-03-05 05:57	wait	25	<ul style="list-style-type: none"><li>• <a href="#">SRP000612.1</a> : MCF10A_TSS</li><li>• <a href="#">SR0002122.1</a> : MCF10A_TSS</li><li>• <a href="#">SRX002786.1</a> : MCF10A_TSS</li><li>• 23 runs</li></ul>
<a href="#">SRA008166.3</a>	DDBJ : UT-MGS_33	✔ DDBJ	2009-03-05 03:08	error	21	<ul style="list-style-type: none"><li>• <a href="#">SRR013851.1</a> :: data-load: run file problems failed, error code 35584, message:/sra/staging/loading/SRR013851/run.sh: line 3: 3601 Segmentation fault /sra/bin/sradb -q -rc /sra/.sradbrc -run SRR013851 -platform ILLUMINA -rps 1 -batch batch.cfg -log run.log</li></ul>
<a href="#">SRA008164.3</a>	DDBJ : UT-MGS_31	✔ DDBJ	2009-03-05 03:07	error	13	<ul style="list-style-type: none"><li>• <a href="#">SRR013813.1</a> :: data-load: run file problems failed, error code 35584, message:/sra/staging/loading/SRR013813/run.sh: line 3: 20572 Segmentation fault /sra/bin/sradb -q -rc /sra/.sradbrc -run SRR013813 -platform ILLUMINA -rps 1 -batch batch.cfg -log run.log</li></ul>

➡ 問題のないオブジェクトに対してアクセッション番号が順に割り振られる

Entrez

Pubmed、GEO、Genome Project、Taxonomy



**Short Read Archive**

Project ID の  
拡張、共通化



生データ

Sequence count 型  
定量データ



**GEO**  
Gene Expression Omnib

加工データ



**INSDC** International Nucleotide Sequence Database Collaboration

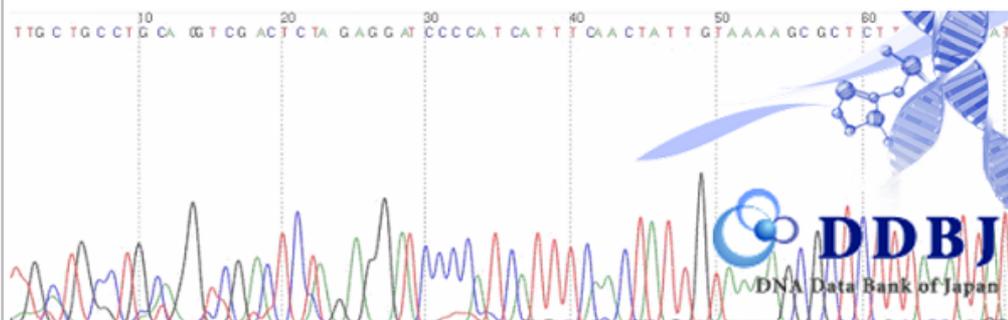
# **DDBJ Read Archive (DRA)**



## 2008年～ DDBJ Read Archive

項番	機関名	生物種	機種	件数	容量 (圧縮時)	SRA発行	作業期間	状況
1	東京大学 新領域 菅野・鈴木研	寄生虫6種 と媒介昆虫 2種	Solexa	1億件	327 GB	2008/ 9/14	約2ヶ月	完了
2	理研 林崎研	ヒト	Solexa	20万件	38 MB (fastq)	2008/ 11/12	約2ヶ月半	完了
3	非公開	非公開	Solexa	200万件	23 GB	2008/ 12/12	約2ヶ月	完了
4	非公開	非公開	Solexa	4000万件	454 GB	2009/ 2/27	約2ヶ月	完了
5	非公開	非公開	454	不明	752 MB	2009/ 1/30	約1ヶ月半	完了
6	非公開	非公開	Solexa	不明	604 GB	2009/ 5/4	約1ヶ月半	完了
7	非公開	非公開	Solexa	不明	2.3 GB (fastq)	2009/ 5/7	約1ヶ月半	NCBI アップロード済
8	非公開	非公開	Solexa	不明	167 GB	未完了	未完了	対応中
9	非公開	非公開	Solexa	不明	3.8 GB (fastq)	2009/ 5/14	約1ヶ月	NCBI アップロード済
10	非公開	非公開	454	不明	3.9 GB	未完了	未完了	対応中
11	非公開	非公開	Solexa	不明	不明	未完了	未完了	対応中

## Trace/Short Read Archive 登録



### Trace Archive

Trace Archive は大規模配列決定プロジェクトに由来する DNA sequence chromatogram (trace), 塩基判定 (base call), 品質評価 (quality estimate) のアーカイブです。 [Trace Archive \(NCBI\)](#) および [Ensembl Trace Server](#) とともにデータベースを共同構築しています。

DDBJ においても Trace Archive データの登録を受け付けています。登録をご希望の場合は、

**trace@ddbj.nig.ac.jp** までご連絡ください。

DDBJ で受け付けたデータは [DDBJ Trace Archive データ検索サイト](#) で検索、閲覧、ダウンロードが可能です。

また、[FTP 公開サイト](#) でデータの一括取得が可能です。

✳ DDBJ Trace Archive は [ライフサイエンス統合データベースプロジェクト](#) のサポートを受けて運営されています。

### Short Read Archive (SRA)

SRA は 次世代型シーケンサ (454, Solexa, SOLID など) の出力データのアーカイブです。 [Short Reads Archive \(NCBI\)](#) および [European Nucleotide Archive - Reads \(EBI\)](#) とともにデータベースを共同構築しています。DDBJ においても SRA データの登録を受け付けています。登録をご希望の場合は、

**trace@ddbj.nig.ac.jp** までご連絡ください。

SRA へのデータ登録には、加工処理 (アダプター配列の除去やフィルタリング等) されていない出力データ (もしくは fastq 形式のデータ) が必要です。登録に必要なファイルに関しては [NCBI SRA Submission Guidelines](#) をご覧ください。

1. SRA 登録受付システムの確立
2. DDBJ 独自のアクセッション番号発行 (DRxxxxxx)
3. SRAデータの利用者への提供

## 増大する Short Read データを DDBJ/EBI/NCBI 3極で協力してアーカイブすべく国際会議が開催された

5月14-15日 at NCBI

中村保一、舘野義男、真島淳、児玉悠一（DDBJ出席者）



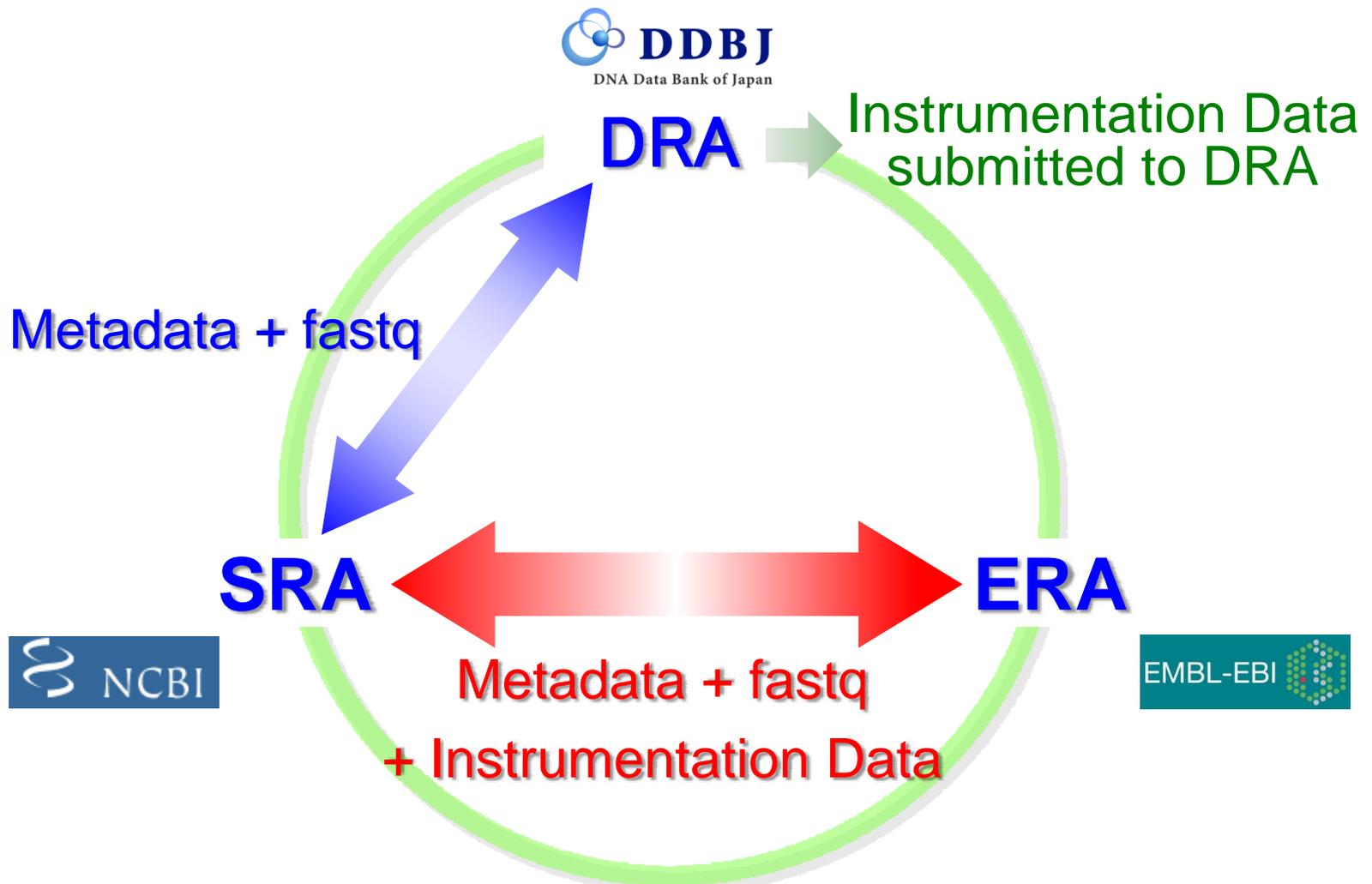
Trace/Short Read Archive Meeting

1. INSDC の枠組みの中に DRA/ERA/SRA を位置づける  
(6月の Advisory Meeting で正式決定される模様)
2. DDBJ が自前のシステムを確立するまでの間、  
仲介登録されたデータに対して NCBI SRA が DRxxxxxx 番号を発行する
3. 3局でメタデータと fastq データを共有する
4. XML メタデータの詳細を定めた “XML Writer’s Guide” を作成する
5. メタデータの Minimum Information を定める
6. 3極で合意を得てから、XML スキーマを変更する
7. “生”データの登録、保存に関するガイドラインを作成する
8. Short Read データに対してもアクセッション番号を要求するよう  
Journal に Letter を送付する
9. 論文中でのデータ引用方法を明確化し、Journal Office に伝える

(重要事項のみ抜粋)

➡ 国際塩基配列データベースと同様の運営を目指すことで合意

# データ交換



現在、NCBI と EBI はすべてのデータを交換している  
DDBJ に登録された”生”データは DDBJ でのみ公開する