

# 古くて新しい課題： 生物情報資源相互運用性の 実現

国立遺伝学研究所

生命情報・DDBJ研究センター

データベース運用開発研究室

菅原秀明

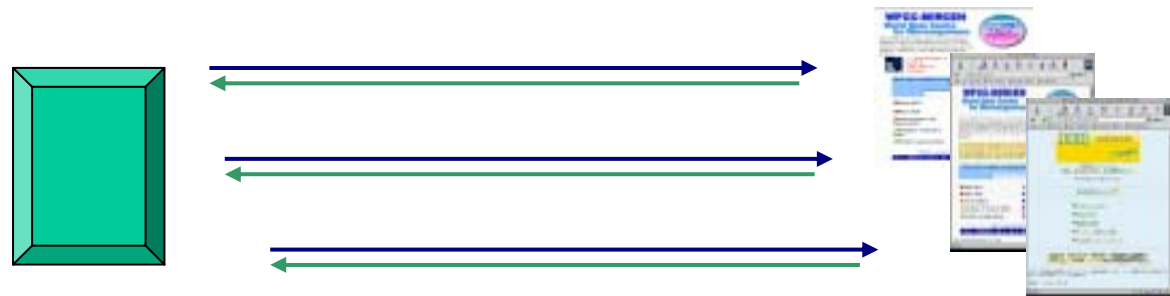


古くて～

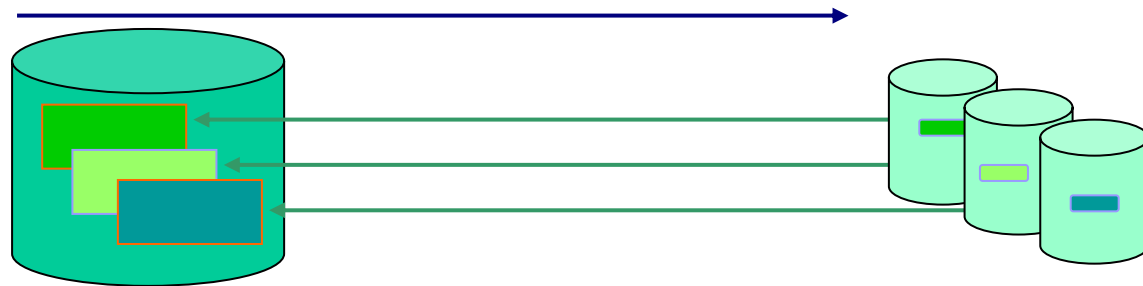
生物情報資源の利用者側  
から

# 分散統合システム

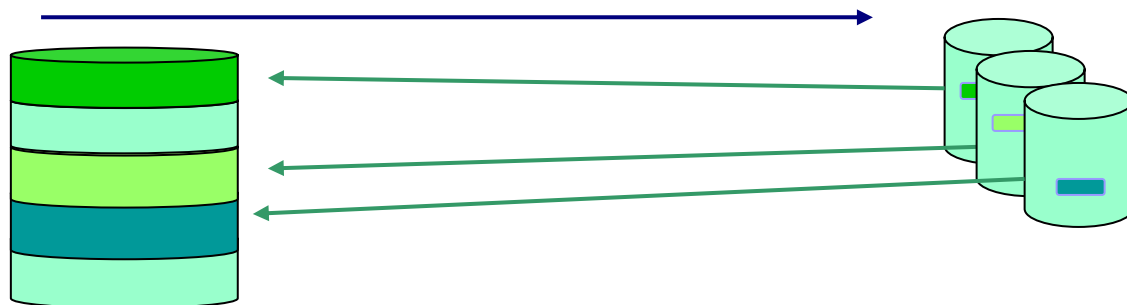
Web 閲覧  
copy & paste



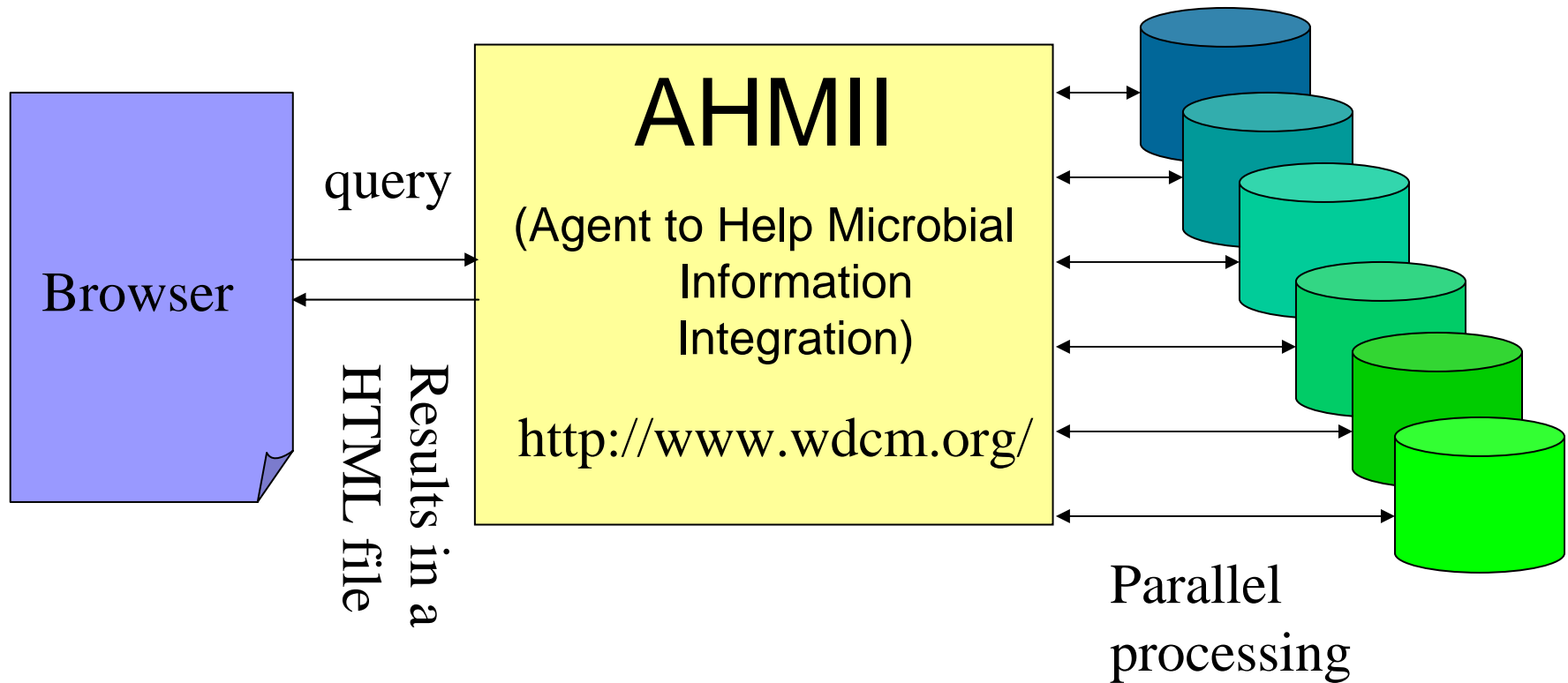
平行検索



構造化



# 平行検索



# 平行検索

*Scheme*

*Bacteria*

*Fungi & Yeasts*

*Cell lines*

## Bacteria Search

### Database:

[Select-All](#) [Clear-All](#)

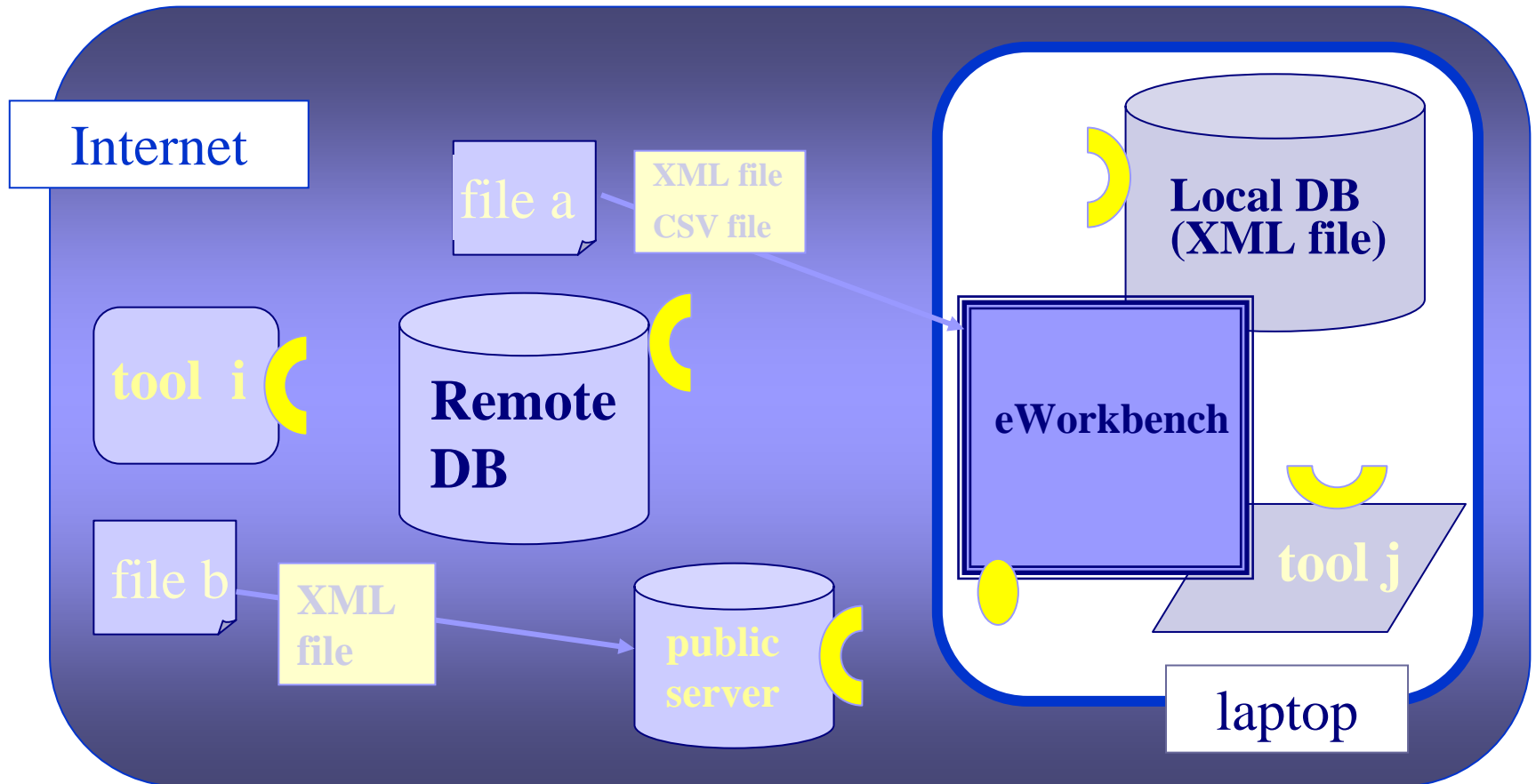
- ATCC** (ATCC Bacteria)
- WDCM** (WDCM STRAINS Bacteria)
- JCM** (Japan Collection of Microorganisms)
- CBS** (CBS Bacteria Database)
- TXSearch** (DDBJ TXSearch for Taxonomy Database)
- BCCM** (BCCM Bacteria Database)
- CICT** (CICT Bacteria NameDatabase)
- ICB** (ICB database by MBI)
- CABRI** (CABRI Bacteria Catalogues)
- IFO** (IFO Bacteria database)
- DSMZ** (DSMZ Bacteria)

### Scientific name:

Genus :

Species :

# CORBAによる構造化



: CORBA (Common Object Request Broker Architecture)

# CORBAによる構造化: Local DB

The screenshot displays the Infor BIO V4.50 software interface. The main window is titled "Infor BIO V4.50 - [C:\Ye-Workbench\SampleData\JCM\_Basidio\_DB\Basidio\_221 strains.xml]". The interface includes a menu bar (Import DB, Search & Edit, Analysis, Option, Web / Help) and a toolbar. The "Search & Edit" window is active, showing search categories and keywords. Below the search fields, there are buttons for "Keywords Search", "Search ALL", "Data Check", "Export DB(Sum File)", "Export DB(Divide Fil...", and "Export Txt File". The "Search Results" section contains a table with the following data:

Category No.	Fixed RID	Fixed INPUT-DATE	Fixed GENUS	Fixed SPECIES	Fixed COLONIES	Fixed STRAIN	Fixed STATUS	Sequence DOBJ ACCESS	Sequence 18S rRNA
00001	Sample 1	2001/02/18							atagctgtgt...
00002	Sample 2	2001/02/18							taccbogggat...
00003	JCM6865	2001/08/22	Bensington...	ciliata		JCM 6865	Phylogene@c	D38233	agtcatatgct...
00004	NZ-3	2001/08/17	Bensington...	ingoldii		JCM 7445T...	Phenotypic/...	D38234	agtcatatgct...
00005	NB-146	2001/08/17	Bensington...	miscanthi		JCM 5733T...			
00006	NB-146	2001/08/17	Bensington...	miscanthi		JCM 5733T...	Phenotypic/...	D38236	agtcatatgct...
00007	K-304	2001/08/07	Bensington...	musae		JCM 8801T...	Phenotypic/...	D43946	aagtcatatgc...
00008	NB-186	2001/08/17	Bensington...	naganoens...	Yellowish b...	JCM 5978T...	Phenotypic/...	D38366	agtcatatgct...
00009	JCM7476	2001/08/22	Bensington...	phyllada		JCM 7476T...	Phylogene@c	D38237	agtcatatgct...
00010	OK-255	2001/08/06	Bensington...	sakaguchii		JCM 10047...	Phenotypic/...	AB001746	agtcatatgct...
00011	NB-150	2001/08/17	Bensington...	subrosea		JCM 5735T...	Phenotypic/...	D38238	agtcatatgct...

The interface also features a sidebar with "e-Workben" branding, a UNESCO logo, and text including "MICROBIAL RESOURCES CENTRE MIRCEA UNESCO" and "All rights reserved". At the bottom, there are buttons for "Select: Row", "1 Entry View", "2 Entry Compare", "New Entry", "Delete Entry", "Import Txt File", and "Edit Mode: OFF". The status bar shows "Ready" and "Look and Feel: Windo...".

# CORBAによる構造化: Remote DB

Infor BIO V4.50 - [C:\e-Workbench\SampleData\FJCM\_Basidio\_DB\FBasidio\_221strains.xml]

Import DB | Search & Edit | Analysis | Option | Web / Help

Local DB | Remote DB

Remote DB

CORBA DB List

Reload DB List

DOBJ-NUC  
DOBJ-DAD

Search Result

Get Schema							
No.	RID	Date	Genus	Species	Strain	DDBJ Acc.	Sequence
00001	AB001746	09-DEC-19...	Bensington...	sakaguchii	OK255 (=J...	AB001746	agtcatatgct
00002	AB001747	09-DEC-19...	Bensington...	sakaguchii	OK259 (=J...	AB001747	agtcatatgct
00003	AB038050	27-FEB-20...	Bensington...	ingoldii	JCM 7445	AB038050	ctgcggagg
00004	AB038051	27-FEB-20...	Bensington...	musae	JCM 8801	AB038051	ctgcggagg
00005	AB038052	27-FEB-20...	Bensington...	sakaguchii	JCM 10047	AB038052	ctgcggagg
00006	AB038061	27-FEB-20...	Bensington...	miscanthi	JCM 5733	AB038061	ctgcggagg
00007	AB038062	27-FEB-20...	Bensington...	phyllada	JCM 7476	AB038062	ctgcggagg
00008	AB038063	27-FEB-20...	Bensington...	subrosea	JCM 5735	AB038063	ctgcggagg
00009	AB038064	27-FEB-20...	Bensington...	yuccicola	JCM 6251	AB038064	ctgcggagg
00010	AB038065	27-FEB-20...	Bensington...	ciliata	JCM 6865	AB038065	ctgcggagg
00011	AB038066	27-FEB-20...	Bensington...	naganoens...	JCM 5978	AB038066	ctgcggagg
00012	AB038079	27-FEB-20...	Bensington...	yamatoana	JCM 3896	AB038079	ctgcggagg
00013	AB040114	20-JUN-20...	Bensington...	thailandica	JCM 10651	AB040114	gctcagtagtc
00014	AB040115	20-JUN-20...	Bensington...	thailandica	JCM 10655	AB040115	tgtgaaactg
00015	AB040116	20-JUN-20...	Bensington...	thailandica	JCM 10657	AB040116	tgtgaaactg
00016	AB040117	20-JUN-20...	Bensington...	thailandica	JCM 10658	AB040117	atcctcctc
00017	AB040647	20-JUN-20...	Bensington...	ciliata	IFM 49224/	AB040647	aaccctctac
00018	AB040648	20-JUN-20...	Bensington...	yuccicola	IFM 49223/	AB040648	aaccctctac
00019	AF101826	04-NOV-19...	Bensington...	yamatoana	AS 2.1973	AF101826	gctgctcctg
00020	AF109987	23-APR-20...	Bensington...	ciliata	CR87514	AF109987	taaccnaac

Search Keywords

Genus:

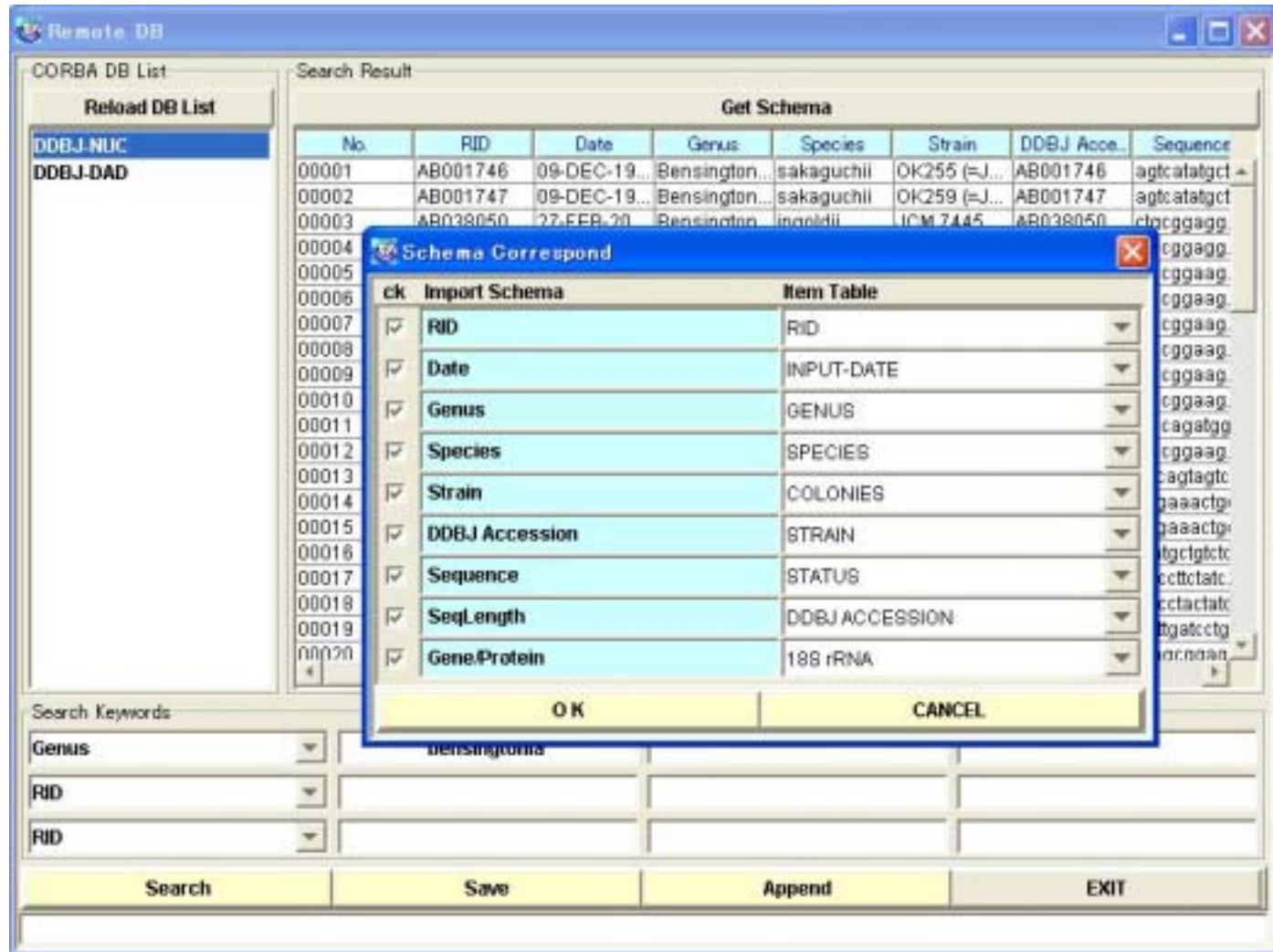
RID:

RID:

Search Save Append EXIT

69 Entry are found!

# CORBAによる構造化：スキーマ照合



# CORBAによる構造化：スキーマ照合

The screenshot shows the 'Remote DB' application interface. On the left, there is a 'CORBA DB List' with 'DDBJ-NUC' and 'DDBJ-DAD' listed. The main area displays a 'Search Result' table with columns: No., RID, Date, Genus, Species, Strain, DDBJ Acce., and Sequence. A 'Schema Correspond' dialog box is overlaid on the table, showing a mapping between the table's columns and a 'ck Import Schema' table. The dialog has an 'OK' button at the bottom.

No.	RID	Date	Genus	Species	Strain	DDBJ Acce.	Sequence
00001	AB001746	09-DEC-19...	Bensington	sakaguchii	OK255 (=J...	AB001746	agtcatatgct
00002	AB001747	09-DEC-19...	Bensington	sakaguchii	OK259 (=J...	AB001747	agtcatatgct
00003	AB038050	27-FEB-20...	Bensington	monidii	JCM 7445	AB038050	ctgcggagg
00004							cggaag
00005							cggaag
00006							cggaag
00007							cggaag
00008							cggaag
00009							cggaag
00010							cggaag
00011							cagatgg
00012							cggaag
00013							cagtagtc
00014							gaaactgt
00015							gaaactgt
00016							tgctgtctc
00017							ctctatc
00018							ctctatc
00019							tgatctg
00020							tcncaag

The 'Schema Correspond' dialog box contains the following table:

ck	Import Schema	Item Table
<input checked="" type="checkbox"/>	RID	RID
<input checked="" type="checkbox"/>	Date	INPUT-DATE
<input checked="" type="checkbox"/>	Genus	GENUS
<input checked="" type="checkbox"/>	Species	SPECIES
<input type="checkbox"/>	Strain	COLONIES
<input checked="" type="checkbox"/>	DDBJ Accession	STRAIN
<input checked="" type="checkbox"/>	Sequence	STATUS
<input checked="" type="checkbox"/>	SeqLength	STATUS
<input checked="" type="checkbox"/>	Gene/Protein	DDBJ ACCESSION

The 'Gene/Protein' dropdown menu is open, showing options: 18S rRNA, ITS, JCM, C. Glucose, C. Galactose, and C. L. Sorbose.



新しい～

生物情報資源の提供側  
から

# Nucleic Acids Research

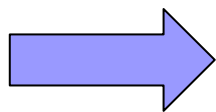
## Database Categories List

[www3.oup.co.uk/nar/database/c/](http://www3.oup.co.uk/nar/database/c/)

- ▶ Major Sequence Repositories
- ▶ Comparative Genomics
- ▶ Gene Expression
- ▶ Gene Identification and Structure
- ▶ Genetic and Physical Maps
- ▶ Genomic Databases
- ▶ Intermolecular Interactions
- ▶ Metabolic Pathways and Cellular Regulation
- ▶ Mutation Databases
- ▶ Pathology
- ▶ Protein Databases
- ▶ Protein Sequence Motifs
- ▶ Proteome Resources
- ▶ RNA Sequences
- ▶ Retrieval Systems and Database Structure
- ▶ Structure
- ▶ Transgenics
- ▶ Varied Biomedical Content

# 公共データベースの役割

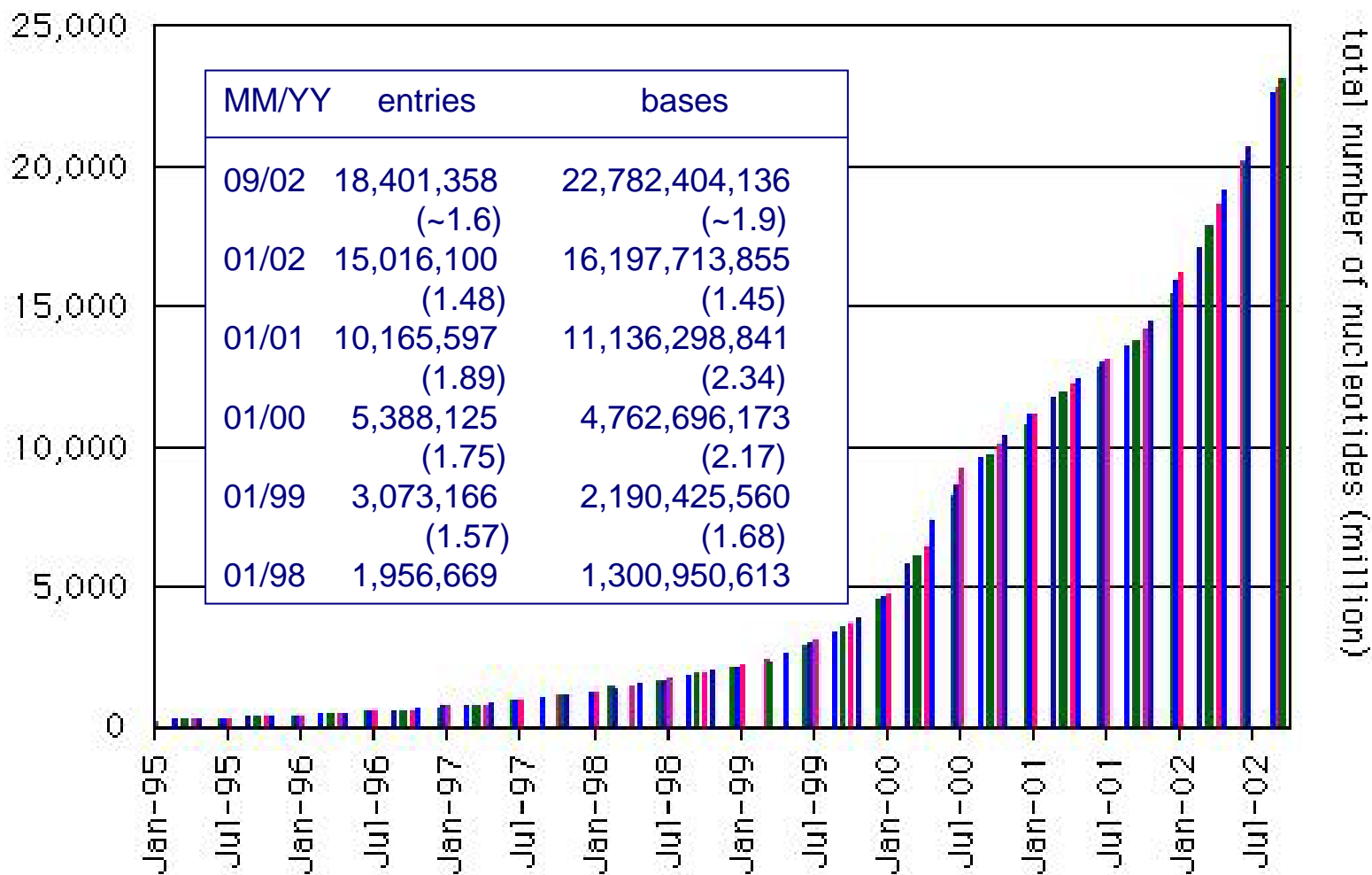
- 網羅的情報を提供する。
- 基準となる情報を提供する。
- 透明な構造で情報を提供する。
- 使いやすい道具を提供する。



DNA Data Bank of Japan  
(DDBJ)の例

# 塩基配列データの津波

DDBJ/EMBL/GenBank database growth 1995/1-2002/9



All right reserved © 国立遺伝学研究所

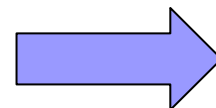
菅原秀明

# ゲノム配列決定や環境からのダイレクトシーケンシングの継続・増加

	US	Europe	Japan	Others
Archaea	11	4	2	1
Procaryote	168	63	12	14
Eucaryote	112	54	6	3

Ref <http://wit.integratedgenomics.com/GOLD/>

Environmental sequences expands in INSD  
e.g. 97,512 entries (September, 2002)  
103,938 entries (October, 2002)  
107,936 entries (November, 2002)



**scalability**

Results from round 2

# INSDの典型的利用例

Sequences producing significant alignments:  
Sequences used in model and found again:

	score	E
	(bits)	Value
dad  <a href="#">AB003309-1</a>  BAA33499.1  367 Helicobacter pylori cdrA pr		
prf  <a href="#">2421352A</a>  cdrA gene		
dad  <a href="#">AE000528-7</a>  AAD07141.1  139 Helicobacter pylori 26695 H		
pir  <a href="#">H64527</a>  H64527 hypothetical protein HP0064 - Helicobacte		
dad  <a href="#">AE001445-6</a>  AAD05643.1  139 Helicobacter pylori J99 put		
pir  <a href="#">G71978</a>  G71978 hypothetical protein jhp0059 - Helicobact		
dad  <a href="#">AX044620-1</a>  CAC17214.1  346 Helicobacter pylori protein		
dad  <a href="#">AE000528-9</a>  AAD07129.1  831 Helicobacter pylori 26695 conse...	74	3e-12
pir  <a href="#">B64528</a>  B64528 conserved hypothetical ATP-binding protein HP...	74	3e-12
dad  <a href="#">AE001445-8</a>  AAD05645.1  806 Helicobacter pylori J99 put	70	5e-11
pir  <a href="#">A71979</a>  A71979 hypothetical protein jhp0061 - Helicobact		

Sequences not found previously or no

dad|[AF007761-1](#)|AAD04237.1| 287|Stre

CONVERGED!

[Go to top](#)

>dad|[AB003309-1](#)|BAA3349  
Length = 367

Score = 310 bits (786)  
Identities = 160/160 (100%), Positives = 160/160 (100%)

## 30万ORF総当り 解析

## 51,000配列の アラインメント

## 500,000 blasts/day in Celera

```

WEFVEPLNEKALSGLEYQLKMGLSDAFKDFVKRSNYGFSQWRSFMVGNK 60
WEFVEPLNEKALSGLEYQLKMGLSDAFKDFVKRSNYGFSQWRSFMVGNK
WEFVEPLNEKALSGLEYQLKMGLSDAFKDFVKRSNYGFSQWRSFMVGNK 60

LDGKGLLIDFMQSLKEWLEPEEIIFANDGYEGYLLNTASDVYVFLDTD 120
LDGKGLLIDFMQSLKEWLEPEEIIFANDGYEGYLLNTASDVYVFLDTD
LDGKGLLIDFMQSLKEWLEPEEIIFANDGYEGYLLNTASDVYVFLDTD 120

```

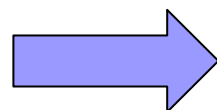
```

Query: 121 DGSRHALLHLKMFLLKRELKDLQREQDFWTKSSQFKYSYP 160
          DGSRHALLHLKMFLLKRELKDLQREQDFWTKSSQFKYSYP
Sbjct: 121 DGSRHALLHLKMFLLKRELKDLQREQDFWTKSSQFKYSYP 160

```

# 津波対策

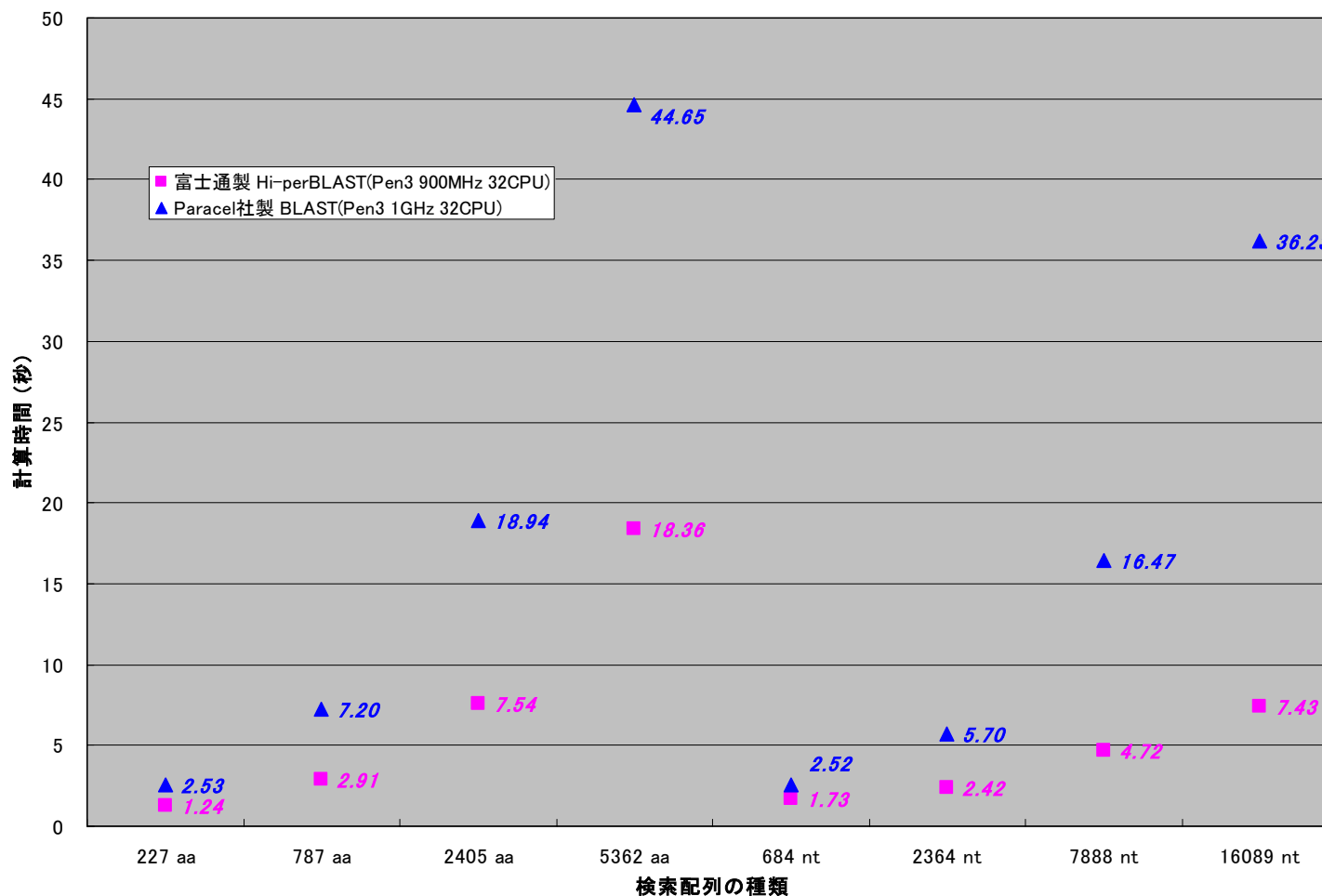
- 種々の検索用インデックス作成
  - 24hrs以内に終了するか?
- DDBJのBLASTサーバーは「入場規制中」



PCクラスターと並列処理  
による解決

# 並列Blastの評価

検索配列が1つの時の性能比較



# 並列マルチプルアライメント評価

500件、平均長1100残基

CPU数	処理時間(s)
1	24310
2	15682
4	8280
8	4623
16	2950

Clustal-G:

Kuo-Bin Li (Bioinformatics Institute, Singapore)

All right reserved © 国立遺伝学研究所

菅原秀明

# 津波対策

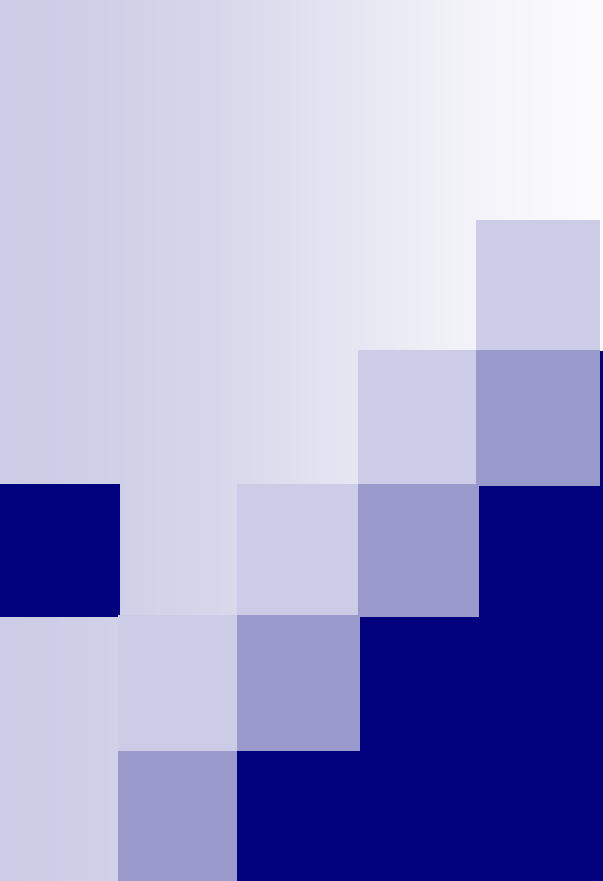
- データバンクの個別対応？
- コミュニティーとしての対応？
  - (e.g. GRID)

# 「透明な構造・使いやすさ」の提供

- 通信・交換プロトコール
- データ・フォーマット
- データ表現
- データ内容(意味)



相互運用性を高める



**XML**

# エントリー例

LOCUS AK025000 2589 bp mRNA HUM 29-SEP-2000  
DEFINITION Homo sapiens cDNA: FLJ21347 fis, clone COL02724.  
ACCESSION AK025000  
VERSION AK025000.1  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens colon cDNA to mRNA, clone\_lib:COL clone:COL02724.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2589)  
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,  
Shibahara,T., Tanaka,T. and Nakamura,Y.  
TITLE Direct Submission  
JOURNAL Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio  
Sugano, Institute of Medical Science, University of Tokyo,  
Laboratory of Genome Structure

FEATURES  
source

Location/Qualifiers

1..2589

/clone="COL02724"

/clone\_lib="COL"

/note="cloning vector pME18SFL3"

/organism="Homo sapiens"

/sequenced\_mol="cDNA to mRNA"

/tissue\_type="colon"

18..2378

/codon\_start=1

/protein\_id="BAB15051.1"

/translation="MLGARAWLGRVLLLPRAGAGLAASRRGSSSRDKDRSATVSSSVP

CDS

SFLSRQLPFLSTLRRLEDQATAYVCENQACSVPI TDPCELRKLLHP"

BASE COUNT 529 a 797 c 773 g 490 t 0 others

ORIGIN

1 atcgccccg agcagccatg ctgggcgcgc gggcctggtt gggccgcgtc cttctgctgc

annotation

protein sequence

DNA sequence

# データ構造

## ■ The DDBJ/EMBL/GenBank **Feature Table**: Definition

- **Feature key**: e.g. *CDS (coding sequence; sequence of nucleotides that corresponds with the sequence of amino acids in a protein; ---)*
- **Qualifier**: e.g. */gene (symbol of the gene corresponding to a sequence region)*
- **Value**, e.g. *“text”*

## ■ Taxonomy Database

- **Taxonomic name**: *Escherichia coli*  
[\[Go to lower taxa\]](#) [\[Get sequence\(s\) related to this taxon\]](#)
- **Full lineage**: [cellular organisms](#); [Bacteria](#); [Proteobacteria](#); [gamma subdivision](#); [Enterobacteriaceae group](#); [Enterobacteriaceae](#); [Escherichia](#)

# エントリーAB000100の例

## FF (Flat File) format

```
CDS      121..912
         /gene="cynB"
         /codon_start=1
         /transl_table=11
         /product="intrinsic membrane protein"
         /protein_id="BAA21794.1"
         /translation="MVRTPVPLYLRWAVSILSVLAFLAIWQIAAASGFLGKTFFGSLR
         TLQDLFGWLSDPFFDNGPNDLGIGWNLLISLRRVAIGYLLATVVAIPLGIAIGMSALA
         -----"
```

## XML document

```
<cds>
  <location>121..912</location>
  <qualifiers name="codon_start">1</qualifiers>
  <qualifiers name="gene">cynB</qualifiers>
  <qualifiers name="product">intrinsic membrane protein</qualifiers>
  <qualifiers name="protein_id">BAA21794.1</qualifiers>
  <qualifiers name="translation">MVRTPVPLYLRWAVSILSVLAFLAIWQIAAASGFLGKTFFG
    SLRTLQDL ----- LLDQGFRLFLENQFSYAGNR</qualifiers>
  <qualifiers name="transl_table">11</qualifiers>
</cds>
```

All right reserved © 国立遺伝学研究所

# XML, XML, XML

BSML (The Bioinformatic Sequence Markup Language)

NCBI-XML (National Center for Biotechnology Information XML)

BSA(Biomolecular Sequence Analysis)

TIGR XML (The Institute Genome Research XML)

XEMBL(EMBL project)

GAME (Genome Annotation Markup Language)

AGAVE (Architecture for Genomic Annotation, Visualization and Exchange)

DAS(Distributed Sequence Annotation System)

GO (Gene Ontology)

# XML, XML, XML

InterPro (composite search system of proteins)

ProML (Protein Markup Language)

PROXIML (PROtein eXtensible Markup Language)

PDBj-ML (Protein Data Bank Japan Markup Language)

MAGE-ML (MicroArray Markup Language)

MaXML (RIKEN mouse cDNAs)

CellML (Cell Markup Language)

PML (Polymorphism Markup Language)

BIND XML (Biomolecular Interaction Network Database)

SBML (System Biology Markup Language)

BIOML (BIOpolymer Markup Language)

CML (Chemical Markup Language)





# SOAP

# SOAP

- Simple Object Access Protocol 1.1  
(W3C Note 08 May 2000)

- Messaging in XML on HTTP

Open and interoperable

(conf. CORBA, RPC, RMI)

Fire-wall friendly (conf. IIOP)

- Independent of platforms and languages

---

RPC: Remote Procedure Call

RMI: Remote Method Invocation

All right reserved © 国立遺伝学研究所

菅原秀明

# データとツールをSOAPで提供

- BLAST (homology search)
- FASTA (homology search)
- SSearch (Smith-Waterman homology search)
- GetEntry (retrieve entries by accession number(s))
- DDBJ (get the DDBJ full entry or extract some Features)
- ClustalW (multiple alignment)
- SRS (Sequence Retrieval System)
- TxSearch (taxonomy database search)



# XML Central of DDBJ

## »» DDBJ-XML **XML**

DDBJ-XML is a new output format of DDBJ entries. It is readable both for human and machine.

## »» Web services **SOAP**

This is the first public SOAP service for biology in Japan.  
The project aims at the standardization of bioinformatics services on the Internet and the improvement of the interoperability. This page also provides you a Web interface of the SOAP server.

## »» Registration/Publication of your Web services

You are courteously invited to register your Web service(s) in the list, if you open bioinformatics SOAP service(s) to the public.

## »» Sample program to bind a web service to your program

You can use a Web service by specifying the URL address, method and parameter(s) in your program such as Java or perl.

## »» Demo system using SOAP technology

You can get the seamless access to the retrieval of DDBJ entry and BLAST execution.  
Enjoy SOAP technology.

## »» SOAP tutorial

It's easy to access the SOAP services.  
This is the first step to try web services.



# WSDL

# WSDL

- Web Services Description Language (WSDL) 1.1 (W3C Note 15 March 2001)
- WSDL is an XML format for describing network services as a set of endpoints operating on messages containing either document-oriented or procedure-oriented information.

# Webサービスのリスト



WSDL List

[How to execute with SOAP](#)  
[How to execute with web I/F](#)  
[WSDL Registrant Menu](#)

## Method

[getFFEntry\(accession\)](#)

[getXMLEntry\(accession\)](#)

[getFeatureInfo\(accession, feature\)](#)

[getAllFeatures\(accession\)](#)

[getRelatedFeatures\(accession, start, stop\)](#)

[getRelatedFeaturesSeq\(accession, start, stop\)](#)

[\[contact us\]](#)

# 例1 アノテーションが付与されている 部分配列を切り出す

[Method]

getRelatedFeatures(accession, start, stop)

[Question]

Retrieve the feature information  
corresponding to the sub-sequence  
between 59000th base and 64000th base  
of AL121903

## [Result]

repeat\_region 423..717

CDS join(37..121,4775..4917)

repeat\_region 1775..2064

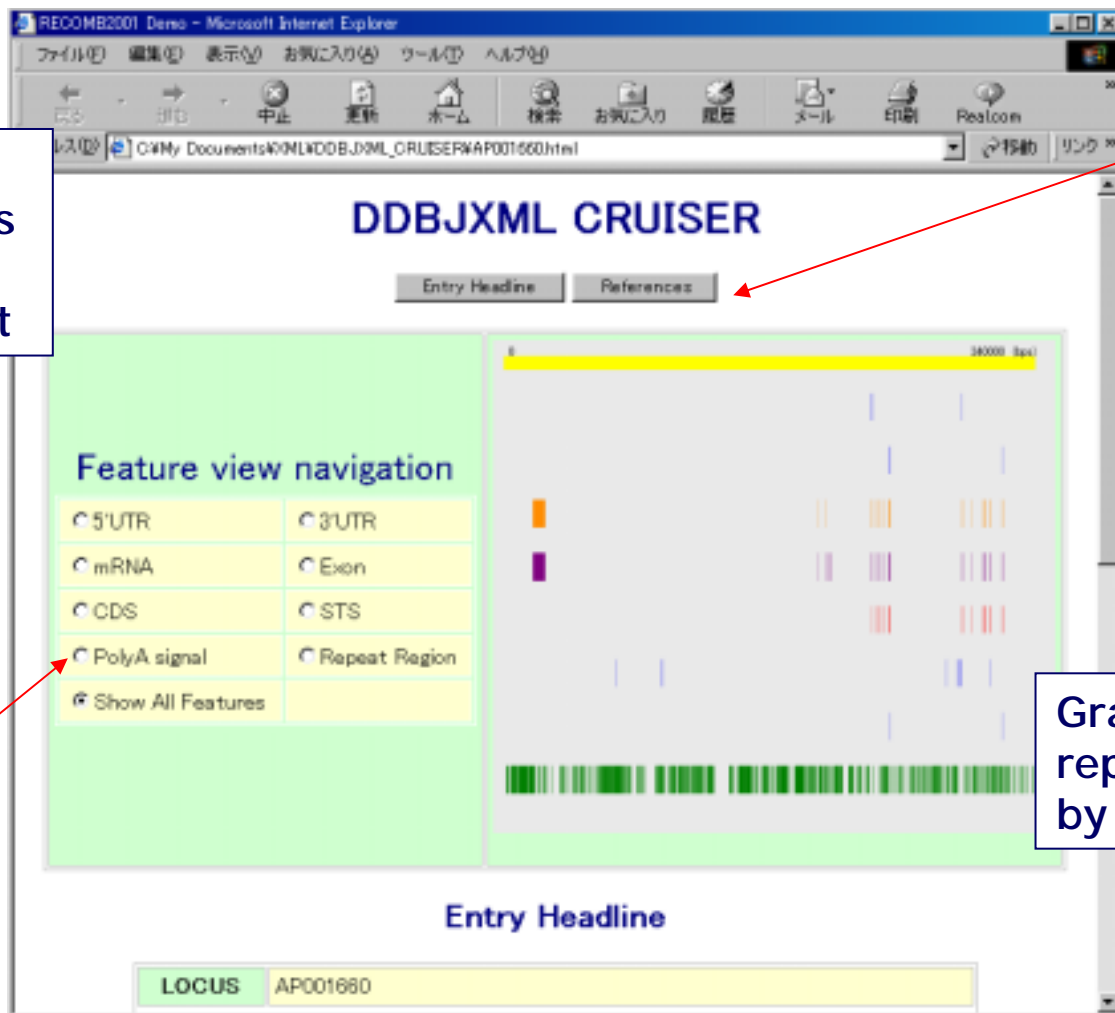
repeat\_region 2067..2362

source 1..5001

repeat\_region 3067..3374

mRNA join(26..121,4775..4917)

# 図示へ



The page structure is created by JAVA script

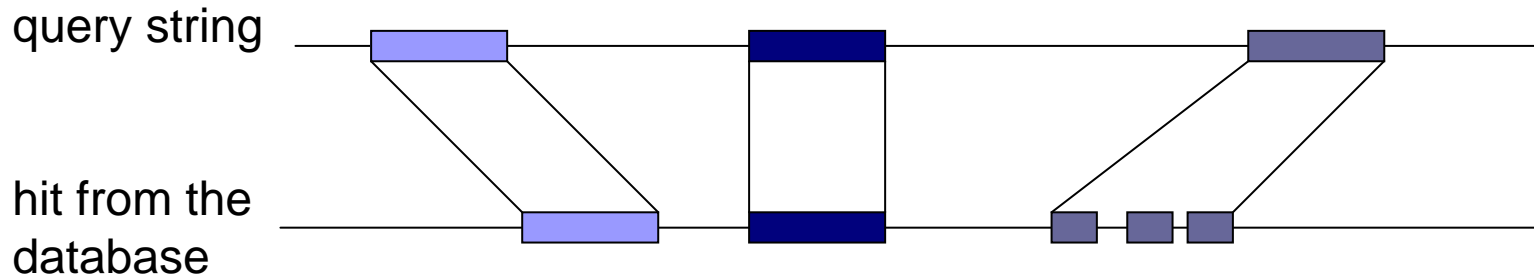
clickable

Graphical representation by SVG

clickable

# 例2 Blastの結果からアライン部分を切り出す

[Method]  
extractPosition(result)



# [Result of extractPosition()]

AF058428 | AF058428.1

Query 86            248

Hit     86            248

Query 320           384

Hit     320           384

Query 564           601

Hit     561           598

. . .

# Webブラウザのインターフェースも用意

<b>ENDPOINT</b>	http://oak.genes.nig.ac.jp/glue/urn:blast
<b>METHOD</b>	search
<b>INPUTS</b>	<b>program</b> <input type="text" value="&lt;string&gt;"/>
	<b>database</b> <input type="text" value="&lt;string&gt;"/>
	<b>query</b> <input type="text" value="&lt;string&gt;"/>
	<b>param</b> <input type="text" value="&lt;string&gt;"/>
<b>OUTPUTS</b>	Result <string>

Execute

# 例3 CDSの塩基配列を切り出す

エントリー(FF)での記述:



欲しいデータ:



**ENDPOINT** `http://xml.nig.ac.jp/xddb/GetEntry`

**METHOD** `getFASTA_CDSEntry`

**INPUTS** **accession**

L17418

**OUTPUTS** **Result** <string>

## Result ( getFASTA\_CDSEntry )

```
>L17418-1|AAB60695.1|7470|join(L17390.1:1921..2041,L17399.1:185..364,L17409.1:135..234,L17419.1
atggggggcctcttctccaagaagccccggagcctgtcgggcccggcgcccgggtctcccc
ttctgctgcggaggatccctgctggcgggttggtgctgcttgcgctgccgggtggcctgg
ggtcaatgcaatgccccagaatggcttccatttgccaggcctaccaacctactgatgaa
tttgagtttcccattgggacatatctgaactatgaatgccgccctggttattccggaaga
ccgttttctatcatctgcctaaaaaactcagtcctggactgggtgctaaggacaggtgcaga
cgtaaatacatgtcgtaatcctccagatcctgtgaatggcatgggtgcatgtgatcaaaggc
atccagttcggatcccaaatataatattcttgtactaaaggataccgactcattggttcc
tcgtctgccacatgcatcatctcaggtgatactgtcatttgggataatgaaacacctatt
tgtgacagaattccttgtgggctacccccaccatcaccaatggagatttcattagcacc
aacagagagaattttactatggatcagtggtgacctaccgctgcaatcctggaagcggg
gggagaaagggtgtttgagcttgtgggtgagccctccatatactgcaccagcaatgacgat
caagtgggcatctggagcggccccgccctcagtgcatatacctaacaaatgcacgcct
ccaaatgtggaaaatggaatattggtatctgacaacagaagcttattttccttaaatgaa
gttgtggagtttaggtgtcagcctggcttctgtcatgaaaggacccccgccgtgtgaagtgc
cagccctgaacaaatgggagccggagctaccaagctgctccagggatgtcagccacct
ccagatgtcctgcatgctgagcgtacccaaagggacaaggacaactttcacctgggcag
gaagtgttctacagctgtgagccccggctacgacctcagaggggctgcgtctatgcgctgc
acaccccagggagactggagccctgcagccccacatgtgaagtgaaatcctgtgatgac
ttcatgggccaacttcttaatggccgtgtgctatttccagtaaatctccagcttggagca
aaagtggattttgtttgtgatgaaggatttcaattaaaaggcagctctgctagtactgt
gtcttggctggaatggaaagcctttggaatagcagtggtccagtggtgtaacaaatctt
```

結合情報

CDS配列データ

# 例4 複数のDBへのアクセス

## BLASTS Demo

This is the DDBJ SOAP Demo system.

Enter the Accession Number and you can get three BLAST results.

-> blastn(DDBJ Bacteria)

GetEntry -> blastx(SwissProt) -> result

-> blastx(PDB)

Accession Number:   (e.g. AF058429)

Retrieve data from the nucleotide sequence database (INSD), the protein sequence database (SWISS-PROT) and the protein 3D structure database (PDB) all together by an accession number (Acc#) referred in a published paper

# Webサービスの簡易登録システム

## WSDL List

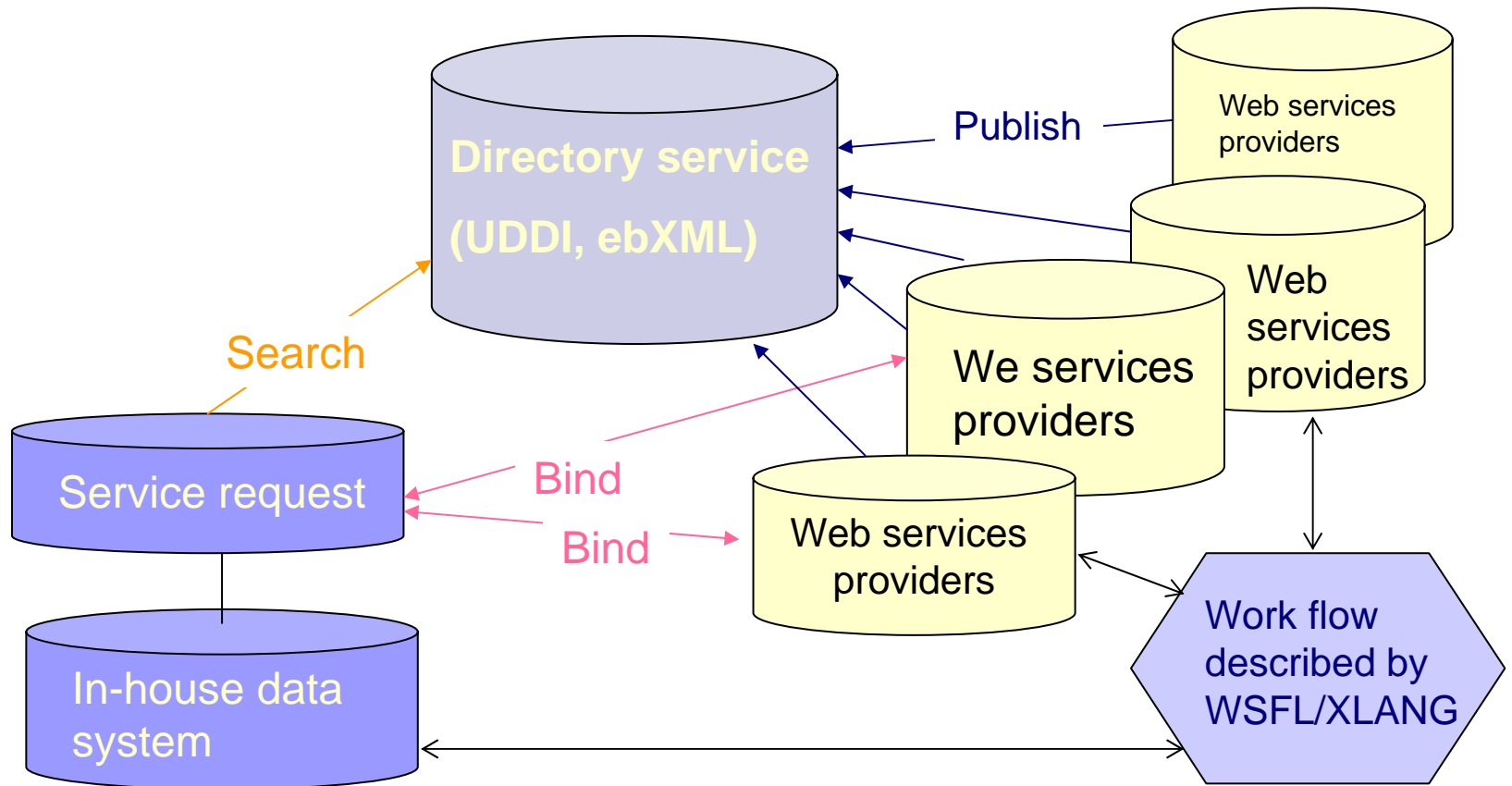
[WSDL Registrant Menu](#)

Name	URL	Document	Registrant	Execute
GetEntry2	<a href="http://xml.nig.ac.jp/wsdl/getentry.wsdl">http://xml.nig.ac.jp/wsdl/getentry.wsdl</a>	<a href="#">document javadoc</a>	<a href="#">Yasumasa Shigemoto2</a>	<input type="button" value="Execute"/>

## WSDL Update

Site Name *	<input type="text" value="GetEntry"/>
WSDL URL *	<input type="text" value="http://xml.nig.ac.jp/wsdl/getentry.wsdl"/>
Document URL	<input type="text" value="http://xml.nig.ac.jp/doc/GetEntry.txt"/>
JavaDoc URL	<input type="text" value="http://xml.nig.ac.jp/javadoc/GetEntry.html"/>

# Web services



Stein, L. (2002) Creating a bioinformatics nation, Nature, **417**, 119-120

All right reserved © 国立遺伝学研究所

菅原秀明

# Web servicesとGRID: Open GRID Service Architecture (GLOBUS 3.0)への期待

Multiple views

Bioinformatics Solution Environment

UDDI (Universal Description, Discovery, and Integration)

WSDL (Web Service Description Language)

SOAP (Simple Object Access Protocol)

GRID

研究者クライアント



Web services



# Resources

# on

# GRID

# align@home

# align@hpcc

# 謝辞

- データベース運用開発研究室  
宮崎 智
- ◆ 富士通株式会社  
山口 政仁  
重元 康正  
松尾 昌嗣