

# *Outline and future plans of KOME & RED databases:*

**Shoshi Kikuchi**

[skikuchi@nias.affrc.go.jp](mailto:skikuchi@nias.affrc.go.jp)

**Plant Genome Research Unit**

**Division of Genome and Biodiversity Research**

**National Institute of Agrobiological Sciences (NIAS)**

# Current status of KOME database

## 1. Global issue

- Rice Full-length cDNA project (Jan.2000~Sept.2003)
- Collaboration of NIAS-FAIS-RIKEN-BRAIN
- Two methods for FL-cDNA library construction
  - Oligo-capping method (Maruyama and Sugano 1994)
  - Biotinylated cap trapper method, (Carninci et al. 2000)
- *Oryza sativa* L ssp. *japonica* cv. Nipponbare
- 580K FL-EST originated from 380K FL-clones  
(DDBJ accession: CI000001–CI778739)
- 32,127 FL-cDNA clones (completely sequenced clones) from 170K FL-clones  
(DDBJ accession: AK58203–AK122186)  
[http://cdna01.dna.affrc.go.jp/cDNA/ANNOTATE/data\\_num.html](http://cdna01.dna.affrc.go.jp/cDNA/ANNOTATE/data_num.html)

# Front page of KOME site



## Knowledge-based Oryza Molecular biological Encyclopedia

At the beginning of 2000, the rice full-length cDNA project was launched. Three research institutes, the National Institute of Agrobiological Sciences (NIAS), the Foundation for Advancement of International Science (FAIS) and the Institute of Physical and Chemical Research (RIKEN), are collaborating with the coordination of BRAIN (Bio-oriented Technology Research Advancement Institute) on this project.



This site is administrated by [National Institute of Agrobiological Sciences](#)

No.116314

Our article is selected as one of "Fast Breaking Papers" by ISI Essential Science Indicator.

KEYWORD

Search

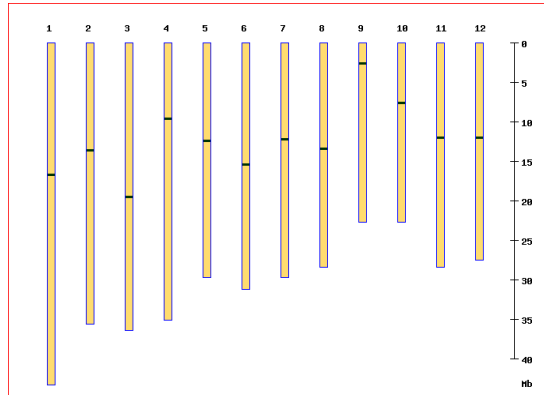
- What's new Update information of this page **NEW**
- About About this Site
- Search clones Search clones by keyword
- TU index TU information based on TIGR pseudomolecules **NEW**
- Accession ID index Accession ID index, Annotation list of Accession ID
- Cluster ID index Cluster ID index, Annotation list of cluster ID
- Mapping viewer Mapping result to the TIGR Pseudomolecules
- Motif index Motif ID index, InterPro Search Result
- Similarity search Similarity search to our cDNA clones
- GO search Search clones by Gene Ontology
- Distribution Distribution of cDNA clones
- Sitemap Site Map
- Link NIAS Rice genome tools
- FTP download FTP Download

**Note:** Netscape Navigator 6.x, Internet Explorer 5.x or upper version recommended (on Windows 98 or Windows 2000).

# Current status of KOME database

## 2. Mapping and alignment to the genomic DNA

- Mapping and alignment to Pseudomolecules
- TIGR rel 3, IRGSP built 3, BGI (93-11)



Number of Unified TUs					
CHR	AE	NAE	ANE	EST	ALL
1	3534	657	3278	87	7556
2	2795	527	2596	72	5990
3	3174	483	2752	58	6467
4	2239	472	3234	40	5985
5	2036	342	2572	36	4986
6	2024	427	2769	46	5266
7	1947	338	2629	38	4952
8	1682	343	2600	37	4662
9	1336	322	2019	29	3706
10	1310	306	2352	20	3988
11	1464	322	2744	32	4562
12	1427	368	2678	24	4497
all	24968	4907	32223	519	62617

**AE:** Annotated Expressed, **NAE:** Non-annotated Expressed, **ANE:** Annotated Non-expressed, **EST:** on 9K Array

Unmapped clones		
TYPE	Number	DETAIL
FL-cDNA	1471	<a href="#">list</a>

# Comparison of FL-EST mapping results among three pseudomolecules

---

	Number of TU determined
TIGR Pseudomolecules rel.3	29,875
IRGSP Pseudomolecules blt.3	29,884
BGI (93-11)	29,734

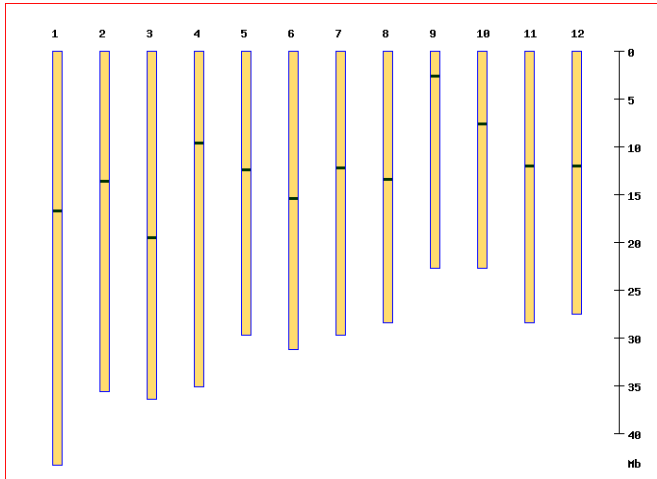
---

**Numbers of TU on *japonica* and *indica* are almost same!**

# Current status of KOME database

## 2. Mapping and alignment to the genomic DNA

- Mapping and alignment to Pseudomolecules
- TIGR rel 3, IRGSP built 3,



Unmapped clones		
TYPE	Number	DETAIL
FL-cDNA	1471	<a href="#">list</a>

Number of Unified TUs					
CHR	AE	NAE	ANE	EST	ALL
1	3534	657	3278	87	7556
2	2795	527	2596	72	5990
3	3174	483	2752	58	6467
4	2239	472	3234	40	5985
5	2036	342	2572	36	4986
6	2024	427	2769	46	5266
7	1947	338	2629	38	4952
8	1682	343	2600	37	4662
9	1336	322	2019	29	3706
10	1310	306	2352	20	3988
11	1464	322	2744	32	4562
12	1427	368	2678	24	4497
all	24968	4907	32223	519	62617

**AE:** Annotated Expressed, **NAE:** Non-annotated Expressed, **ANE:** Annotated Non-expressed, **EST:** on 9K Array

# Detailed information of TU

Start and End Nt  
in TIGR's Pm

FL-cDNA

FL-EST

CDS by TIGR

[PREV 1 Mb](#)

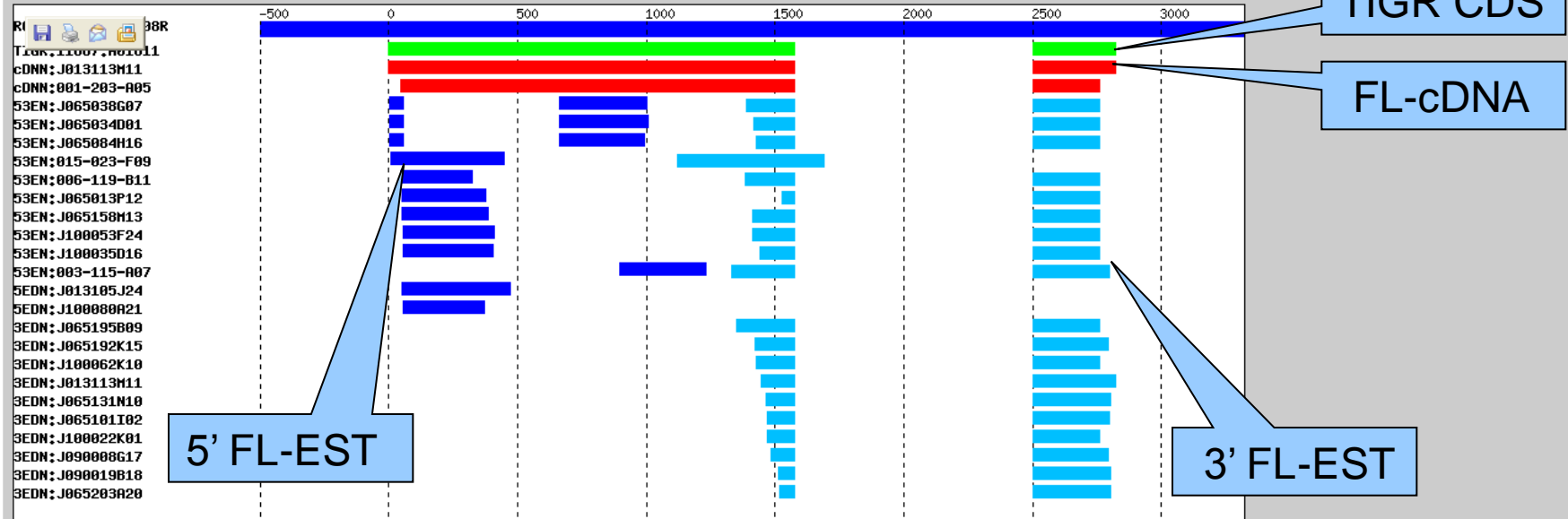
[NEXT 1 Mb](#)

UNIFIED TU ID	TU ID	CHR	START	END	LENGTH	STRAND	cDNA	END	CDS	TYPE
R01-016170-1A	<a href="#">R01-Cv3DPA-090005F</a>	1	9009766	9013493	3728	1	001-029-B02	both53end	11667.t01502	AE
R01-016180-2A	<a href="#">R01-Cv3DPA-090006F</a>	1	9014925	9016539	1615	1		both53end	11667.t01503	AE
R01-016190-1A	<a href="#">R01-Cv3DPA-090003R</a>	1	9016622	9020830	4209	-1	J013098K24	both53end	11667.t01504	AE
R01-016200-1AR	<a href="#">R01-Cv3DPA-090004R</a>	1	9021680	9025656	3977	-1	J023053G08	both53end	11667.t01505	AE
R01-016210-3A	<a href="#">R01-Cv3DPA-090005R</a>	1	9026523	9031838	5316	-1		3end	11667.t01506	AE
R01-016220-1A	<a href="#">R01-Cv3DPA-090006R</a>	1	9037567	9050732	13166	-1	J013127H23,J033081B1...	both53end	11667.t01507	AE
R01-016230-5A		1	9050293	9050710	418	1			11667.t01508	ANE
R01-016240-1A	<a href="#">R01-Cv3DPA-090007R</a>	1	9054616	9057841	3226	-1	002-180-E02	both53end	11667.t01509	AE
R01-016250-1AR	<a href="#">R01-Cv3DPA-090008R</a>	1	9060749	9063568	2820	-1	001-203-A05,J013113M...	both53end	11667.t01510	AE
R01-016260-2A	<a href="#">R01-Cv3DPA-090009R</a>	1	9064144	9068178	4035	-1		both53end	11667.t01511	AE
R01-016270-1A	<a href="#">R01-Cv3DPA-090007F</a>	1	9071394	9080576	9183	1	J033070C16	3end	11667.t01512	AE
R01-016280-1A	<a href="#">R01-Cv3DPA-090008F</a>	1	9084869	9085795	927	1	001-112-G10	both53end	11667.t01513	AE
R01-016290-1A	<a href="#">R01-Cv3DPA-090009F</a>	1	9091954	9096294	4341	1	J013044I07	both53end	11667.t01514	AE
R01-016300-1A	<a href="#">R01-Cv3DPA-090010F</a>	1	9096815	9098382	1568	1	001-021-C08	3end	11667.t01514	AE
R01-016310-3A	<a href="#">R01-Cv3DPA-090011F</a>	1	9098383	9099501	1119	1		3end	11667.t01514	AE
R01-016320-1A	<a href="#">R01-Cv3DPA-091006F</a>	1	9110468	9111214	747	1	J023085H22	both53end	11667.t01515	AE
R01-016330-1A	<a href="#">R01-Cv3DPA-091005R</a>	1	9111358	9114206	2849	-1	J033024I08	both53end	11667.t01516	AE
R01-016340-1A	<a href="#">R01-Cv3DPA-091006R</a>	1	9114682	9118193	3512	-1	001-015-A12,J023030N...	both53end	11667.t01516	AE
R01-016350-5		1	9123266	9128614	5349	-1			11667.t01517	ANE
R01-016360-1A	<a href="#">R01-Cv3DPA-091007R</a>	1	9130521	9137212	6692	-1	J023019E21	both53end	11667.t01518	AE
R01-016370-1A	<a href="#">R01-Cv3DPA-091008R</a>	1	9156418	9165855	9438	-1	001-103-F09	both53end		NAE
R01-016380-1A	<a href="#">R01-Cv3DPA-091007F</a>	1	9156534	9162468	5935	1	J013001G08	both53end	11667.t01519	AE
R01-016390-1AR	<a href="#">R01-Cv3DPA-091008F</a>	1	9163134	9166753	3620	1	J013022M19	both53end	11667.t01520	AE
R01-016400-5A		1	9168410	9169681	1272	1			11667.t01521	ANE
R01-016410-2A	<a href="#">R01-Cv3DPA-091009R</a>	1	9169957	9172681	2725	-1		both53end	11667.t01522	AE
R01-016420-4R		1	9169960	9170249	290	1				ANE
R01-016430-4R		1	9170258	9172517	2260	1				ANE
R01-016440-3A	<a href="#">R01-Cv3DPA-091010R</a>	1	9175997	9181291	5295	-1		3end	11667.t01523	AE
R01-016450-5A		1	9189031	9190221	1191	1			11667.t01524	ANE
R01-016460-2AR	<a href="#">R01-Cv3DPA-091011R</a>	1	9192308	9195751	3444	-1		both53end	11667.t01525	AE

# TU report page

## TU REPORT

TU Report, R01-Cv3DPA-090008R (9060749bp-9063568bp, Chr:1)



On rare occasions the start site of alignment between cDNA and genomic DNA sequences does not fit into the "0" point. It is because the "0" point, which has been determined according to the cDNA's mapping onto the genomic DNA sequence by BLAT program, has not been followed by sim4 program used for the graphical indication.

### TU Detail

TU name	R01-Cv3DPA-090008R
Position on Chr.	090 kbp, Chr. 1 (9060749bp-9063568bp)
32k FL-cDNA (Red)	J013113M11:(PLACE), 001-203-A05:(PLACE)
TIGR CDS (Green): <i>link to TIGR</i>	11667.m01510 (11667.m01611)
5End OPSs (Blue)	003-M115F-A07(PLACE), 006-M119F-B11(PLACE), 015-M023F-F09(PLACE), J01B3105J24T3(PLACE), J06B5013P12T3(PLACE), J06B5034D01T3(PLACE), J06B5038G07T3(PLACE), J06B5084H16T3(PLACE), J06B5158M13T3(PLACE), J10B0035D16T3(PLACE), J10B0053F24T3(PLACE), J10B0080A21T3(PLACE)
3End OPSs (Skyblue)	003-M115R-A07, 006-M119R-B11, 015-M023R-F09, J01B3113M11M3, J06B5013P12M3, J06B5034D01M3, J06B5038G07M3, J06B5084H16M3, J06B5101I02M3, J06B5131N10M3, J06B5158M13M3, J06B5192K15M3, J06B5195B09M3, J06B5203A20M3, J09B0008G17M3, J09B0019B18M3, J10B0022K01M3, J10B0035D16M3, J10B0053F24M3, J10B0062K10M3
Rice PIPELINE	<input type="text" value="R01-Cv3DPA-090008R"/>
Annotation from FL-cDNA	Arabidopsis thaliana At3g56860.T8M16_190 mRNA, complete cds. PLN Arabidopsis thaliana At3g56860.T8M16_190 mRNA, complete cds. PLN

Link to PLACE

# OPS (one pass sequence) report page

## OPS Detail

Clone name	015-023-F09
5' OPS name	015-M023F-F09
TU ID (5' OPS)	R01-Cv3DPA-090008R
5' Duplicate TU	
3' OPS name	015-M023R-F09
TU ID (3' OPS)	R01-Cv3DPA-090008R
3' Duplicate TU	
PLACE ( plant cis-element database)	R01-Cv3DPA-090008R
Rice PIPELINE	<input type="button" value="015-M023R-F09"/> <input type="button" value="015-M023F-F09"/>

## Sequence

### Nucleotide Sequence(015-M023R-F09)

```
ATTTCTTTACCGATGTAATAAATGCAAGATCAATTTCTTGGCAGAAAAACACATGTGAAAACTATTTGTAGGTTCCCTATGGTTGGGACACAATCACCTATC  
CGGATCTTACCTAATGACCCATGTAATTTCTTCTGGGCTCCTGGTGGAGTACTGCCATAACCACCACCACCCCATAGCCGCCAAATAACCAACCAGC  
ACCAGGCATCCCAGGCAAAACCACTACTACCCAGTCTGCAGAAAGCCCTGGGTTAGGCCAACCCCTGAACATTTGCACCAACTCCCAGGATGTTGTTTCAGA  
CCCAACCCTCCACCCTGGGAGGCAAGGATAGCTGTCAAGGCCTGCCCCAGAGCAGGATTCACCCCCTGGGCCAACAGCTACACCTGGCAATGAAAGTCAAAAC  
TGGATACTGGAGATGGGCATCACATGGCCCCAACAGCAGCACCCAGGCAAAAGAAATGACTATGAGTCCATAACCACCAGCGCCTTCTCCCGCTGAAAT  
GCCAGCACCATACAAAACCCCAAGTCTCCACCTTTGTTGGGCTTGGTCCATCAATTGCCTTCTGGC
```

### Nucleotide Sequence(015-M023F-F09)

```
AGACTCCACACCACACAGCCCCCTTTCCOCTCCCTCCACAGCCGTCGAAAACCCCTAGCCCCAAGGTTCCCCCAGCGCTCGCGCTCGCCGTCGCCATGGG  
CAAGAAGGGGAAGCTCGATTCCAAATCCCCCGCCGCGCTAGTTCGCGAGCCGCGAGGSCCGCAGCAGCCGCGAGCCGCGCCGCGCCGCGCCGCGCGT  
GCCGAGCCCTCTCCAGCCCGAAGCCCTAGCCGAGGACCCGSCCCCTCGTCCGAGCCCTGGGCCTATCCAGCGAGGGCGCGGGGAGAGGATGATGT  
CCCGTGAAGCTGGCGGTGGCGAGGAAAGAGGAGGTGGAGGAGGTCGAGGTGGAGGAGGAGGTGGAGGTGACGAGGACGAGGATGGGAGGGGGAGGGGA  
GGAGGAGGAGGAGGCGGCGGAGAGGACGCCACTCGATCC
```

[BACK](#) [HOME](#)

# KOME report page: Information of FL-cDNA

[\[Clone Detail\]](#) [\[Nucleotide Analysis\]](#) [\[Amino acid Analysis\]](#) [\[Sequences\]](#)

[HELP](#)

## Clone Detail

[HELP](#)

Clone name	J013113M11 <input type="button" value="PIPELINE"/>
Data source	RIKEN
Clone ID	210732
Accession NO	<a href="#">AK099896</a>
Length (bp)	1895
Sequence Quality (%)	100
Sequence Accuracy (%)	100
Cluster ID	13530
Cluster member	<a href="#">001-203-A05[1786bp]</a>
TU ID (TIGR Pseudomolecule ver3.0)	<a href="#">R01-Cv3DPA-090008R</a> (9060749 bp - 9063568 bp Chr.1)
Locus ID	<a href="#">R01-AP001080-40060R</a> (BAC: <a href="#">AP001080</a> 91975 bp - 94794 bp Chr.1)
Same Locus ID	<a href="#">001-203-A05</a>
Library infomation	shoot

## Nucleotide Analysis

[HELP](#)

GenBank hit (by BLASTN)	DEFINITION	ACCESSION NO	SCORE	EXPECT		
GenBank hit (by BLASTX) <input type="button" value="result"/>	DEFINITION <b>Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, clone:P0499C11.1PL11</b>	<a href="#">AP001080.1</a>	573	1e-162		
Genome annotation (from GenBank hit)	DEFINITION <b>Oryza sativa genomic DNA, chromosome 4, BAC clone: OSJIBb0034H3, complete sequence.</b>	<a href="#">AL606657.4</a>	241	8e-62		
Product annotation (from GenBank hit)	DEFINITION <b>Arabidopsis thaliana AT3g56860/T8M16_190 mRNA, complete cds.</b>	<a href="#">AF367280.1</a>	295	4e-78		
<i>Oryza sativa ssp. Japonica</i> hit (by BLASTN)	DEFINITION <b>AP001080 Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, clone:P0499C11</b>	<a href="#">AP001080</a>	2264	0.0		
<i>Oryza sativa ssp. indica</i> hit (by BLASTN)	DEFINITION <b>2003-10-07 BGI</b>	<a href="#">Chr01</a>	1142	0.0	STRAND -1	
Rice8987array(g_array) hit (by BLASTN) <input type="button" value="result"/>	ELEMENT NO	CLONE NAME	ACCESSION NO	SCORE	EXPECT	<input type="button" value="Search RED"/>
	7465	S20671_1A	AU056545	1134	0.0	
flank17 (mutant panel databse)	flanking sequence					
PLACE (plant cis-element databse)	<input type="button" value="PLACE result"/>					

## Amino acid Analysis

[HELP](#)

## Promoter Detail

Clone name	015-M023F-F09
TU ID	R01-Cv3DPA-090008R 9060749bp-9063568bp, Chr:1
2kbp upstream element	9063559bp-9065558bp, Strand:-1, Chr:1

## PLACE Result

SIGNAL NAME	SIGNAL ID	STRAND	LOCATION	POS. on Chr	KEYWORD
RYREPEATLEGUMINBOX	S000100	+	23	9065536	RY repeat; legumin box; seed; storage protein;
RYREPEATBNNAPA	S000264	+	23	9065536	RY repeat; RY/G box; seed; napA; napin;
DOFCOREZM	S000265	-	29	9065530	Dof; C4PEPC; CyPPDK; PEP; C4; leaf; shoot;
GATABOX	S000039	-	47	9065512	ASF-2; GATA box; Cab; chlorophyll a b binding protein; leaf;shoot;
INRNTPSADB	S000395	+	51	9065508	initiator; light-responsive transcription; TATA-less promoter;psaDb; lnr element;
POLLEN1LELAT52	S000245	-	55	9065504	pollen; lat52; endo-beta-mannanase; MAI;
TATABOX2	S000109	-	67	9065492	TATA; legA; phaseolin;
WBOXHVIS01	S000442	+	82	9065477	sugar; SURE; patatin; WRKY; isoamylase; SUSIBA2;
MYBCORE	S000176	+	85	9065474	MYB; myb; dehydration; water; stress; flavonoid biosynthesis;leaf; shoot;
POLASIG3	S000088	+	90	9065469	poly A; polyadenylation;
SEF4MOTIFGM7S	S000103	+	94	9065465	soybean; seed; storage protein; 7S; globulin; beta-conglycinin;SEF;
MYBPZM	S000179	+	103	9065456	P; P gene; P-gene; MYB; myb; seed;
REALPHALGLHCB21	S000362	+	105	9065454	REalpha; Lheb21; phytochrome; REbeta;
DOFCOREZM	S000265	+	109	9065450	Dof; C4PEPC; CyPPDK; PEP; C4; leaf; shoot;
MYB1AT	S000408	-	116	9065443	MYB; rd22BP1; ABA; leaf; seed; stress;
POLLEN1LELAT52	S000245	-	119	9065440	pollen; lat52; endo-beta-mannanase; MAI;
BOXINTPATPB	S000296	-	120	9065439	plastid; HEP; atpB; PatpB; HcII; Box I; Box II;
GATABOX	S000039	-	125	9065434	ASF-2; GATA box; Cab; chlorophyll a b binding protein; leaf;shoot;
ROOTMOTIFTAPOX1	S000098	+	144	9065415	root; roId;
GT1CONSENSUS	S000198	-	146	9065413	GT-1; light; TATA; TFIIA; TBP; HR; SAR; TMV; leaf; shoot;
GT1CONSENSUS	S000198	-	147	9065412	GT-1; light; TATA; TFIIA; TBP; HR; SAR; TMV; leaf; shoot;
DOFCOREZM	S000265	-	172	9065387	Dof; C4PEPC; CyPPDK; PEP; C4; leaf; shoot;
ASF1MOTIFCAMV	S000024	+	183	9065376	TGACG; root; leaf; CaMV; 35S; promoter; auxin; salicylic acid;light; as-1; TGA1a, TGA1b; CREB; ASF1; TGA6; shoot; xenobioticstress; SAR; SA; Disease resistance;
HEXMOTIFTAH3H4	S000053	-	183	9065376	hexamer; HBP-1A; HBP-1B; histone H3; CaMV; 35S; IOS; HBP-1;Leucine zipper motif; meristem;
TGACGTVMAMY	S000377	+	183	9065376	alpha-Amylase; cotyledon; seed germination; seed;
ABRELATERD1	S000414	+	185	9065374	ABRE; etiolation; erd;
ACGTATERD1	S000415	+	185	9065374	ACGT; etiolation; erd;
ACGTATERD1	S000415	-	185	9065374	ACGT; etiolation; erd;
DOFCOREZM	S000265	-	193	9065366	Dof; C4PEPC; CyPPDK; PEP; C4; leaf; shoot;
TAAAGSTKST1	S000387	-	193	9065366	KST1; Dof; guard cell;
GT1CONSENSUS	S000198	-	194	9065365	GT-1; light; TATA; TFIIA; TBP; HR; SAR; TMV; leaf; shoot;
IBOXCORE	S000199	-	195	9065364	I box; I-box; rbcS; light regulation; light; leaf; shoot;
GATABOX	S000039	-	196	9065363	ASF-2; GATA box; Cab; chlorophyll a b binding protein; leaf;shoot;
MYBST1	S000180	-	196	9065363	MYB; myb; Myb;
TATCCAOAMY	S000403	+	196	9065363	alpha-amylase; MYB proteins; gibberellin; GA; sugar starvation;
DPBFCOREDCDC3	S000292	-	206	9065353	Dc3; lea class gene; embryo; ABA; DPBF-1, DPBF-2; bZIP; GAI1;ABI5; seed;
DOFCOREZM	S000265	-	214	9065345	Dof; C4PEPC; CyPPDK; PEP; C4; leaf; shoot;
EBOXBNNAPA	S000144	+	224	9065335	napA; storage protein; ABRE; E-box; seed;
EBOXBNNAPA	S000144	-	224	9065335	napA; storage protein; ABRE; E-box; seed;
MYBCORE	S000176	-	224	9065335	MYB; myb; dehydration; water; stress; flavonoid biosynthesis;leaf; shoot;
MYCCONSENSUSAT	S000407	+	224	9065335	MYC; rd22BP1; ABA; leaf; seed; stress; CBF3; cold; CBF.DREB1;ICE1;
MYCCONSENSUSAT	S000407	-	224	9065335	MYC; rd22BP1; ABA; leaf; seed; stress; CBF3; cold; CBF.DREB1;ICE1;
MYB2CONSENSUSAT	S000409	+	224	9065335	MYB; rd22BP1; ABA; leaf; seed; stress;
GATABOX	S000039	-	233	9065326	ASF-2; GATA box; Cab; chlorophyll a b binding protein; leaf;shoot;
SEF4MOTIFGM7S	S000103	+	244	9065315	soybean; seed; storage protein; 7S; globulin; beta-conglycinin;SEF;
GT1CORE	S000125	-	248	9065311	rbcS; box II; GT-1; rbcS-3; leaf; shoot;
MYBCORE	S000176	-	249	9065310	MYB; myb; dehydration; water; stress; flavonoid biosynthesis;leaf; shoot;
RAV1AAT	S000314	-	298	9065261	RAV1; AP2; VP1; B3; root; leaf; shoot;
DOFCOREZM	S000265	-	307	9065252	Dof; C4PEPC; CyPPDK; PEP; C4; leaf; shoot;
POLLEN1LELAT52	S000245	-	313	9065246	pollen; lat52; endo-beta-mannanase; MAI;

Click "PLACE", then list of cis-elements in the 2kb upstream sequence will be shown but ::::

Pipeline system for promoter Analysis will be facilitated.

# Current status of KOME database

## 3. Characteristics of FL-cDNA clones by homology search

- BLAST N and X homology searches to GenBank
- Homology with gene in Arabidopsis
- Product annotation based on BLASTX search
- pSORT: cellular location
- MEMSAT: trans-membrane analyses
- GO classification: process, function and cellular localization from GO attached to the Arabidopsis gene and GO attached to InterPro motif

Kikuchi et al. Science (2003) 301, 261-420

# Detail of KOME report page

GenBank hit (by BLASTX) <input type="button" value="result"/>	<b>DEFINITION</b> Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, clone:P0499C11,PLII	<b>ACCESSION NO</b> AP001080.1	<b>SCORE</b> 573	<b>EXPECT</b> 1e-162	
Genome annotation (from GenBank hit)	<b>DEFINITION</b> Oryza sativa genomic DNA, chromosome 4, BAC clone: OSJNBb0034H13, complete sequence.	<b>ACCESSION NO</b> AL606657.4	<b>SCORE</b> 241	<b>EXPECT</b> 8e-62	
Product annotation (from GenBank hit)	<b>DEFINITION</b> Arabidopsis thaliana AT3g56860.T8M16_190 mRNA, complete cds.	<b>ACCESSION NO</b> AF367280.1	<b>SCORE</b> 295	<b>EXPECT</b> 4e-78	
Oryza sativa ssp. Japonica hit (by BLASTN)	<b>DEFINITION</b> AP001080 Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, clone:P0499C11	<b>ACCESSION NO</b> AP001080	<b>SCORE</b> 2264	<b>EXPECT</b> 0.0	
Oryza sativa ssp. indica hit (by BLASTN)	<b>DEFINITION</b> 2003-10-07 BGI	<b>ACCESSION NO</b> Chr01	<b>SCORE</b> 1142	<b>EXPECT</b> 0.0	<b>STRAND</b> -1
Rice8987array(g_array) hit (by BLASTN) <input type="button" value="result"/>	<b>ELEMENT NO</b> <b>CLONE NAME</b> <b>ACCESSION NO</b> <b>SCORE</b> <b>EXPECT</b>	<input type="button" value="Search RED"/>			
	7465    S20671_1A    AU056545    1134    0.0				
Top17 (mutant panel databse)	flanking sequence				
PLACE (plant cis-element databse)	<input type="button" value="PLACE result"/>				

## Amino acid Analysis

[HELP](#)

ORF(Longest)	106 bp - 1578 bp (77.73 %)				
Length(aa)	490				
Arabidopsis hit (by BLASTP)	<b>DEFINITION</b> UBP1 interacting protein 2a (UBA2a) identical to UBPF1 interacting protein 2a [Arabidopsis thaliana] GI:19682816; contains InterPro entry IPR000504: RNA-binding region RBP-1 (RNA recognition motif) (RRM)	<b>AGILENOCODE</b> At3g56860.3	<b>SCORE</b> 295	<b>EXPECT</b> 4e-80	
PIR (by BLASTX) <input type="button" value="result"/>	<b>DEFINITION</b> hypothetical protein T8M16_190 - Arabidopsis thaliana	<b>PIR ACCESSION</b> P1:T51274	<b>SCORE</b> 295	<b>EXPECT</b> 4e-79	
SWISS-PROT (by BLASTX) <input type="button" value="result"/>	<b>DEFINITION</b> (O92006) RNA-binding protein Musashi homolog 2 (Musashi-2)	<b>SWISS-PROT ENTRY NAME</b> MSI2H_MOUSE	<b>SCORE</b> 80.5	<b>EXPECT</b> 2e-14	
InterPro domain search	<input type="button" value="InterPro result"/>				
BLOCKS search	<input type="button" value="BLOCKS(LongestORF)"/>				
GO classification (GenBank)					
GO classification (InterPro)					
GO classification (Arabidopsis thaliana)	<a href="#">[go:9738]</a> abscisic acid mediated signaling <a href="#">[go:3723]</a> RNA binding <a href="#">[go:3676]</a> nucleic acid binding				
Transmembrane Topology (by MEMSAT)	2(out)				
protein localization (by PROSITE)	chloroplast stroma				
Rice Proteome Database	<b>RPD ID</b>	<b>SCORE</b>	<b>COVERAGE</b>		

# Pipeline system for rice gene search

- **<< Nucleotide sequence query >>**
  - Full-length cDNA (blastn)
  - Full-length cDNA Longest ORF (blastx)
  - Rice8987 Array EST (blastn)
  - japonica* Genome (blastn)
  - indica* Genome (blastn)
  - Mitochondria (blastn)
  - Chloroplast (blastn)
  - TIGR pseudomolecules (blastn)
- **<< Protein sequence query >>**
  - Full-length cDNA (tblastn)
  - Select the target Database

Yazaki et al. Nuc Acids Res 2004 32, D383-D387

# Rice Gene Search

[Japanese](#)

Databases	<b>&lt;&lt; Nucleotide sequence query &gt;&gt;</b> <input checked="" type="checkbox"/> Full-length cDNA (blastn) <input checked="" type="checkbox"/> Full-length cDNA Longest ORF (blastx) <input checked="" type="checkbox"/> Rice8987 Array EST (blastn) <input checked="" type="checkbox"/> Japonica Genome (blastn) <input checked="" type="checkbox"/> Indica Genome (blastn) <input checked="" type="checkbox"/> Mitochondria (blastn) <input checked="" type="checkbox"/> Chloroplast (blastn) <input checked="" type="checkbox"/> TIGR pseudomolecules (blastn)
	<b>&lt;&lt; Protein sequence query &gt;&gt;</b> <input type="checkbox"/> Full-length cDNA (tblastn) Select the target <a href="#">Database</a>
E-VALUE	<input type="text" value="1.0e-10"/> Set the threshold value
Number of descriptions	<input type="text" value="100"/>
Number of alignments	<input type="text" value="100"/>
Filter	<input checked="" type="radio"/> ON <input type="radio"/> OFF Filtering your query sequence for low compositional complexity regions if you select "ON". ( <a href="#">DUST</a> with blastn, <a href="#">SEG</a> with tblastn)
Word Size	<input type="text"/> (default : blastn=11, tblastn=3)
Graphical view	<input checked="" type="checkbox"/> ON (See search result by the Graphical view.)
BLAST original view	<input type="checkbox"/> ON (See search result by the BLAST original format.)
Input sequence	<pre>CTCTTTAACATGCTATATAGATTAGTAGTTAACATGCACCTTCCTTAGT TCAGAGGTATCTTCATTCTAATGCTCATTTATACTAGTAAATGACTGT TAATAATTTTTGTGCCAACCAAAAGATGTGGTTTCTATATCTGCTAAAT CTAACTATATTTCCCGATTCTTAAAATGGCACACTTTTGTGTTGATGAC GTGTTCCCTTTATCCAAGATCTAGTGTTCTTTTGGTACAACTGACATAT CGAATCCCGTTTTTAACCGGCTTAGTTCCTATGCACGGAGAACCTCAT AACTCTGCACATTTTGTGAACTCCTTTGTTTTCTAAGGACGCCCCCACTT GAACTAGCTAACTAAGTTGATTTTAAAGTTTTTTTGGAACTTGGAAAGCAT GTACTAGTCAAGTTTAAAGTGTTTAAATGTGATGGATTGATGGAATCCAAA TGTTCAGTTCAAAATAGAAAATGCTTAAACATGACTTTCGTCACTCATT TCTTAACTCTCGTGAATGATTGCAGGGGGGGTGGGGGGGTGCATGTTAG ACTGTAGTTGAGTAGTTGATTTTGCCTCGGCTTACCGTATTGAACTATA GCAATTGTTCTACCTTGTTTAATGTTTTAGTATGGACAAAATTCACAAA CTTGATATTACAGCTGGAAAATAAATTTTCAAGTGTATCGTGTGAATC GGGTTCCACTCATTGTGAAAATATTTACATATGGGCTAGCATCATTTT</pre>

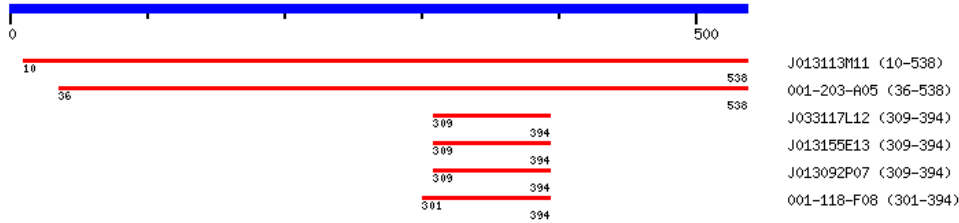
# Results of rice gene search (PIPELINE page)

## WEB Multiple Blast Search Result

### Full-length cDNA Blastn Result

Rank	Clone Name	Cluster ID	Origin	Score	Expect	Library	Mapping Data	Blastn Search Result	Blastx Search Result	Japonica	Indica	Array	TOS17	Identities
1	<a href="#">J013113M11</a>	<a href="#">13530</a>	RIKEN	525	0.0	shoot	<a href="#">AP001080</a> Chr.1		<a href="#">AP001080.1</a>	O	O	O	X	100
2	<a href="#">001-203-A05</a>	<a href="#">13530</a>	FAIS	499	0.0	etiolated shoot	<a href="#">AP001080</a> Chr.1	<a href="#">AK109487.1</a>	<a href="#">AP001080.1</a>	O	O	O	X	100
3	<a href="#">J033117L12</a>	<a href="#">3778</a>	RIKEN	44	1e-16	callus	<a href="#">AP004646</a> Chr.8	<a href="#">AK103059.1</a>		O	O	O	X	90
4	<a href="#">J013155E13</a>	<a href="#">3778</a>	RIKEN	44	1e-16	shoot	<a href="#">AP004646</a> Chr.8			O	O	O	X	90
5	<a href="#">J013092P07</a>	<a href="#">3778</a>	RIKEN	44	1e-16	shoot	<a href="#">AP004646</a> Chr.8	<a href="#">AK071991.1</a>		O	O	O	X	90
6	<a href="#">001-118-F08</a>	<a href="#">12346</a>	FAIS	44	1e-16	UVC irradiated shoot	<a href="#">AP006237</a> Chr.1			O	O	O	X	88

user\_sequence (538bp)



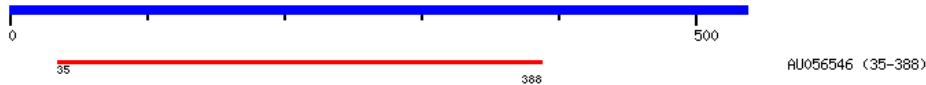
### Full-length cDNA Longest ORF Blastx Result

Rank	Clone Name	Cluster ID	Origin	Score	Expect	Library	Mapping Data	Blastn Search Result	Blastx Search Result	Japonica	Indica	Array	TOS17	Identities
------	------------	------------	--------	-------	--------	---------	--------------	----------------------	----------------------	----------	--------	-------	-------	------------

### Microarray EST Blastn Result

Rank	Array Element	Accession No.	Clone Name	Score	Expect	MAFF	RED	Identities
1	7465	<a href="#">AU056546</a>	S20671_22	346	0.0	<a href="#">MAFF Search</a>	<a href="#">RED</a>	100

user\_sequence (538bp)



### Japonica Genome Blastn Result

Rank	Accession No.	Score	Expect	Description	Identities
------	---------------	-------	--------	-------------	------------

# Future plan of update ~Sept.2006

- 2000 newly sequenced FL-cDNA clones (FY2005~2007, 7000 clones will be completely sequenced) will be added.
- For good support for new oligoarray system (eg. 44K oligoarray)
- Gene family analyses data (part will be shown in Dr. Nagata's talk) such as Ca-related protein, trans-membrane protein, transcription factors will be added.

# Current status of RED

Rice Expression Database

- Microarray project (1999~2003)
- Platform: 1265 and 8987 cDNA based array
- Probe: Full-insert of cDNA clones was stumped (ESTs collected by Sasaki et al. in the early phase of RGP(1991~1997))
- Physiologies from 64 indep research groups
- ca 1300 hybridization data are stored.

# Description of data

- Plant materials  
Germplasm study (Exp factor and Value)
- Treatment  
Condition of treatment, dosage and time course
- Dissected part for sample preparation  
Plant Ontology
- RNA preparation  
Methods, total or polyA RNA, check of quality and quantity
- Hybridization  
Experimental conditions
- Gene expression data  
Normalization method, absolute value, ratio

# Front page of RED



Search RED  
RED Information  
RED top

## Welcome to the Rice Expression Database

### Search contents

- ▶ Data List
- ▶ Search Experiments by Research ID
- ▶ Search Experiments
- ▶ Search Genes
- ▶ Array BLAST Rice 9000

### RED Information

The **Rice Expression Database (RED)** is administrated by, Gene expression team in National Institute of Agrobiological Sciences (NIAS) and Microarray group in Society for Techno-innovation of Agriculture, Forestry and Fisheries (STAFF). RED holds raw and normalized data from expression profiles obtained by the Rice Microarray Project and other research groups. These data are open to the public less than one year after sending the data to each research group. Please refer to the Rice Microarray Opening Site ( **RMOS**: <http://cdna01.dna.affrc.go.jp/RMOS/> ) for full details of our research activities.

[...more RED Information](#)

### References

Yazaki J, Kishimoto N, Ishikawa M, Endo D, Kojima K, MicroArray Center, Kikuchi S. The Rice Expression Database (RED): gateway to rice functional genomics. Trends in Plant Science (2002) Dec 7 (12):563-564

### NIAS Rice genome tools

- ▶  Rice Expression Database (RED)
- ▶  Rice Microarray Opening Site (RMOS)
- ▶  Knowledge-based Oryza Molecular biological Encyclopedia Rice full-length cDNA Database (KOME)
- ▶  Rice Genome Integrated Map Database (INE)
- ▶  Rice Mutant Panel Database (Tost17)
- ▶  Rice Genome Annotation Database (RiceGAAS)
- ▶  Plant Cis-Element motifsearch Database
- ▶ Proteome Database
- ▶  DNA Bank

This page is designed for Internet Explorer 5.0 or later.  
WEB master : [kome\\_web@nias.affrc.go.jp](mailto:kome_web@nias.affrc.go.jp)

# Project members and their research thema



Search RED  
RED Information  
RED top

## Research Groups & Research Subjects

### DATA IN 2001

Research ID	Research Subject	Institute	Section	EMAIL
1000	Development of the rice integral expression monitoring system	Society for Techno-innovation of Agriculture, Forestry and Fisheries (STAFF-Institute)	Group 1	
2101	Monitoring the expression pattern of rice genes under drought, salt and cold stresses	Japan International Research Center for Agricultural Sciences	Biological Resources Division	<a href="mailto:kazukoys@jircas.affrc.go.jp">kazukoys@jircas.affrc.go.jp</a>
2102	A large-scale monitoring of the genes regulated by the treatment with salt in rice seedlings.	National Institute of Agrobiological Resources	Plant Physiology, Disease Physiology	<a href="mailto:tanakayo@abr.affrc.go.jp">tanakayo@abr.affrc.go.jp</a>
2103	DNA microarray analysis for the salt tolerance of rice	Taisei Research Institute	<a href="#">Biotechnology Section Civil &amp; Environment Research Department</a>	<a href="mailto:noboru.endoh@sakura.taisei.co.jp">noboru.endoh@sakura.taisei.co.jp</a>
2104	The analysis of the cascade of gene expression relevant to photoperiodism	National Agricultural Research center	Laboratory of Rice Genetic Engineering	
2105	Expression profiling of UVB- and gamma-ray-irradiated rice plants	National Institute of Agrobiological Science	Laboratory of Gene Expression, Department of Molecular Genetics	<a href="mailto:naoki@abr.affrc.go.jp">naoki@abr.affrc.go.jp</a>
2106	Mobile genetic elements induced by 5-azacytidine treatment	Institute of Molecular and Cellular Biosciences,	<a href="#">Department of Chemical Biology : Biophysics Division</a>	<a href="mailto:hohtsubo@ims.u-tokyo.ac.jp">hohtsubo@ims.u-tokyo.ac.jp</a>
2107	Identification of genes differentially expressed between cultivated and wild rice varieties	National institute for basic biology	Division of gene expression and regulation I	<a href="mailto:shigiida@nibb.ac.jp">shigiida@nibb.ac.jp</a>
2108	Identification and analysis of the genes associated with cold acclimation in wheat	Hokkaido National Agricultural Experiment Station	Department of Low-temperature Sciences,	<a href="mailto:etsuo@cryo.affrc.go.jp">etsuo@cryo.affrc.go.jp</a>
2109	Rice genes that relate aptitude for	Akita Prefectural	<a href="#">Faculty of Bioresource Sciences,</a>	<a href="mailto:konishi@agri.akita-pu.ac.jp">konishi@agri.akita-pu.ac.jp</a>

# Search experiment by Research ID



Search RED  
RED Information  
RED top

## Search Experiments by Research ID

Please select Research ID

► Research ID

- ALL
- 1000
- 2101
- 2102
- 2103
- 2104
- 2105
- 2106
- 2107
- 2108

Display 10 results.(fast)

WEB master : [red@nias.affrc.go.jp](mailto:red@nias.affrc.go.jp)

# In case of 2105, totally 76 experiments are registered.



## Corresponding experimental list

1 - 10 of 76

(Next 10 results)

Experimental ID	Physiology1 of control sample	Physiology2 of control sample
	Physiology1 of target sample	Physiology2 of target sample
2105_0_C0001_T0004	Oxidative	UV, gamma-ray
	Oxidative	UV, gamma-ray
2105_0_C0301_T0304	Oxidative	UV, gamma-ray
	Oxidative	UV, gamma-ray
2105_1_C0001_T0001	Oxidative	UV, gamma-ray
	Oxidative	UV, gamma-ray
2105_1_C0001_T0002	Oxidative	UV, gamma-ray
	Oxidative	UV, gamma-ray
2105_1_C0001_T0003	Oxidative	UV, gamma-ray
	Oxidative	UV, gamma-ray
2105_1_C0001_T0004	Oxidative	UV, gamma-ray
	Oxidative	UV, gamma-ray
2105_1_C0001_T0009	Oxidative	UV, gamma-ray
	Oxidative	UV, gamma-ray
2105_1_C0002_T0001	Oxidative	UV, gamma-ray
	Oxidative	UV, gamma-ray
2105_1_C0002_T0002	Oxidative	UV, gamma-ray
	Oxidative	UV, gamma-ray
2105_1_C0002_T0003	Oxidative	UV, gamma-ray
	Oxidative	UV, gamma-ray

(Next 10 results)

# Global description of the experimental condition

## Experimenter Information

Project ID	2105
Experimental ID	2105_1_C0001_T0003
Experiment title	Expression profiling of UVB- and gamma-ray-irradiated rice plants
User name	Kishimoto.Naoki
Institute	National Institute of Agrobiological Science
Section	Laboratory of Gene Expression, Department of Molecular Genetics

## Common Information

Round	1
Array	Rice8987Arrayf
Basic data	<ul style="list-style-type: none"><li>• <a href="#">histgram</a></li><li>• <a href="#">scatter plot</a></li><li>• <a href="#">signals in spots' order</a></li><li>• <a href="#">picture (1st of C01)</a></li><li>• <a href="#">picture (2nd of C01)</a></li><li>• <a href="#">picture (1st of T03)</a></li><li>• <a href="#">picture (2nd of T03)</a></li></ul>

## Control Data (C0001)

Organism_C	Oryza sativa (rice)
Cultivar_C	Nipponbare
Physiology1_C	Oxidative
Physiology2_C	UV, gamma-ray
stage_C_Keyword	seedling
tissue_C_Keyword	seedling
treatment_C_Keyword	Control
More detail_C	<a href="#">stage,tissue,treatment detail description</a>
total or mRNA_C	t

## Target Data (T0003)

Organism_T	Oryza sativa (rice)
Cultivar_T	Nipponbare
Physiology1_T	Oxidative
Physiology2_T	UV, gamma-ray
stage_T_Keyword	seedling
tissue_T_Keyword	seedling
treatment_T_Keyword	UV
More detail_T	<a href="#">stage,tissue,treatment detail description</a>
total or mRNA_T	t

# Detailed description of the experimental condition



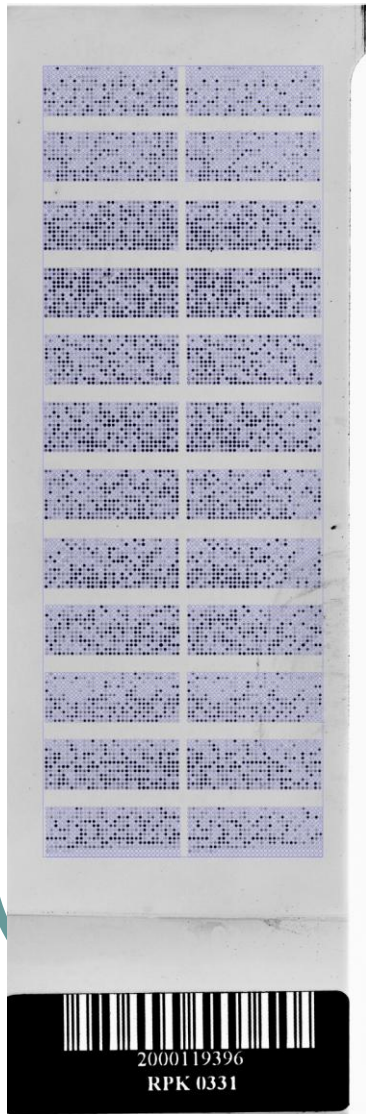
- Search RED
- RED Information
- RED top

## Detail description of 2105\_1\_C0001\_T0003

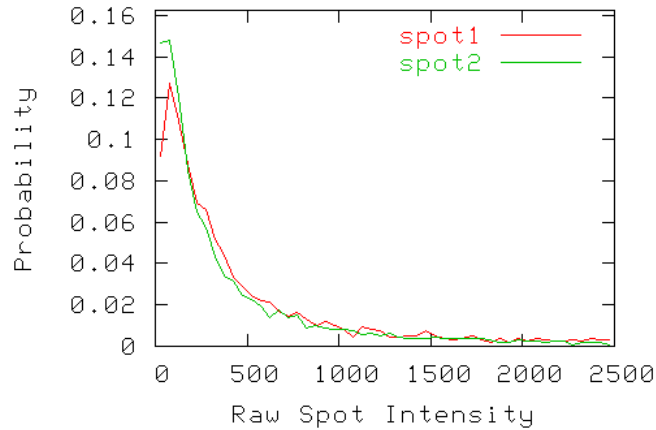
Tissue_stage_C detail	14day-old seedlings (except root); water culture using Hyponex 10-3-3 (1/1000 dilution); 28 degrees C, 16 hours light, 8 hours dark
Treatment_C detail	1 day after nonirradiation (= after same actions without UVB irradiation)

*WEB master : red@nias.affrc.go.jp*

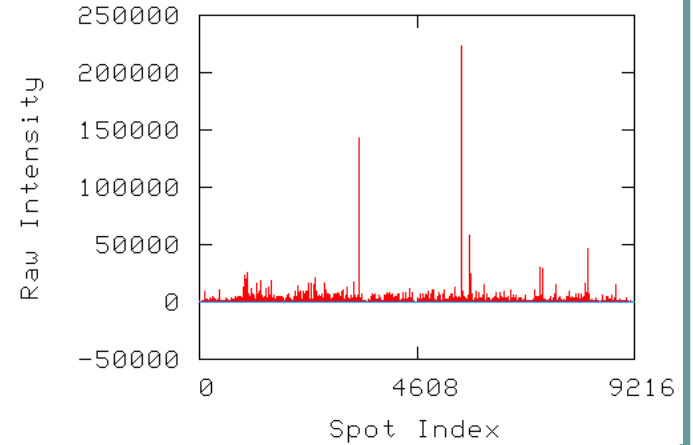
# Examples of data



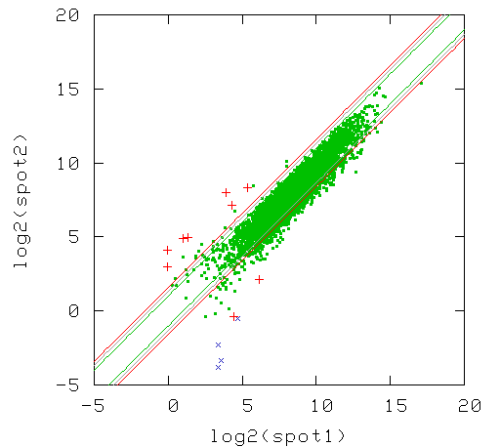
2105 1 C01: Histogram (1/2)



2105 1 C01: Spot 1



2105 1 C01: Spot 1 vs 2 (1/2)



# Rice oligoarray system

- Using 29,100 full-length cDNA sequences
- 60-mer probes: designed by the Agilent Technologies
- considering the  $T_m$  and GC content and removing the possibility of cross-hybridization.
- After several validation experiments using customly-prepared arrays and RNAs from seed, callus, seedlings, etc., a final set of probe sequences was fixed.
- In November 2003, the 22K rice oligomicroarray version 1 (G4138A) was commercialized by the Agilent Technologies, and is now being used by rice molecular biologists world-wide.
- NCBI-GEO platform ID: GPL477 custom GPL892 commercial G4138A  
<http://www.ncbi.nlm.nih.gov/projects/geo/query/browse.cgi?mode=foundplatform>
- Updating from 22K array to 44K array with the information of the mapping results of FL-EST and predicted genes “e-array system”

# Future plan of update

~Sept.2006

- Data using 22K oligoarray will be stored.  
~300 hybridization data
- Wrapped by GCP domain model
- Ontology management
- Load data with MIAME/Plant standard
- Connection to BIO-MOBY system
- Enriched functional gene annotation along with the update of KOME
- Equipment of data mining support tool: Pipeline system for cis-element analysis

# Contributors and collaborators

- NIAS
  - Masaru Takeya (BIO-MOBY)
  - Kouji Satoh (Data prod)
  - Koji Doi (cis-PIPELINE)
  - Toshifumi Nagata (Gene family)
  - Aeni Sasaki-Hosaka
  - Shigemi Iizumi
  - Setsuko Kimura
- Hitachi Software Eng.
  - Kohji Suzuki (WEB)
- IRRI
  - Hei Leung (SP-2 Data prod)
  - Richard Bruskiewich (Domain model)
- CIRAD (France)
  - Pietro Piffanelli (Data prod)
- Nanjin Agric Univ. (China)
  - Guohua Xu
- NIAB (Korea)
  - Ung-Han Yoon (Data prod)
  - Jung-Sook Lee (Data prod)
- POSTEC (Korea)
  - G An (Data prod)
- Gyeongsang National University (Korea)
  - Chang-deok Han (Data prod)
  - Chang-Hyo Goh (Data prod)

These works are supported by MAFF (Rice Genome Projects), BRAIN (Full-length cDNA project) and GCP (SP-2, SP-4)