

# Collection and Comparative Analysis of 1888 Full-length cDNAs from Wild Rice *Oryza rufipogon* Griff. W1943

T LU<sup>1,2</sup>, S U<sup>2,3</sup>, D FAN<sup>2</sup>, J MU<sup>2</sup>, SHANGGUAN<sup>2</sup>, ANG<sup>4</sup>, MINOBE<sup>4</sup>,  
LIN<sup>1</sup>, B HAN<sup>2,\*</sup>

College of Life Science and Biotechnology, Shanghai Jiaotong University, Shanghai, PR China<sup>1</sup>; National Center for Gene Research and Shanghai Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 500 Caobao Road, Shanghai 200233, PR China<sup>2</sup>; School of Life Sciences, Fudan University, Shanghai, PR China<sup>3</sup> and Plant Genome Center, 1-25-2 Kan-nondai, Tsukuba, Ibaraki 305-0856, Japan<sup>4</sup>

(Received 29 August 2008; Accepted 9 July 2008; Available online 6 August 2008)

## Abstract

A huge amount of cDNA and EST resources have been developed for cultivated rice species *Oryza sativa*; however, only few cDNA resources are available for wild rice species. In this study, we isolated and completely sequenced 1888 putative full-length cDNA (FLcDNA) clones from wild rice *Oryza rufipogon* Griff. W1943 for comparative analysis between wild and cultivated rice species. Two cDNA libraries were constructed from 3-week-old leaf samples under either normal or cold-treated conditions. Homology searching of these cDNA sequences revealed that >96.8% of the wild rice cDNAs were matched to the cultivated rice *O. sativa* ssp. *japonica* cv. Nipponbare genome sequence. However, <22% of them were fully matched to the cv. Nipponbare genome sequence. The comparative analysis showed that *O. rufipogon* W1943 had greater similarity to *O. sativa* ssp. *japonica* than to ssp. *indica* cultivars. In addition, 17 novel rice cDNAs were identified, and 41 putative tissue-specific expression genes were defined through searching the rice massively parallel signature-sequencing database. In conclusion, these FLcDNA clones are a resource for further function verification and could be broadly utilized in rice biological studies.

**Key words:** ; *Oryza rufipogon*; - DNA; ; -

## 1. Introduction

*Oryza rufipogon* Griff. (AA genome) is a wild rice species that is closely related to *O. sativa*.

21

\* Corresponding author. E-mail: hanb@sjb.ac.cn. Tel: +86 21-64845260. Fax: +86 21-64825775. E-mail: hanb@sjb.ac.cn

C DNA  
 -  
 .<sup>10</sup> RNA , FL DNA ,  
 minuta (BBCC 5211 EST O.  
 Oryza rufipogon ).<sup>11</sup>  
 .<sup>12</sup> 1943 O. rufi-  
 pogon. F 1888 FL DNA O.  
 rufipogon 1943  
 ; (>96.8%)  
 1943 . F ,  
 . indica. A , 1% . japonica  
 1943 FL DNA  
 . 41 -  
 - (MPSS) .<sup>13</sup>

## 2. Materials and Methods

### 2.1. Plant materials and cDNA library construction

T FL DNA  
 O. rufipogon G . 1943. S -  
 / 13/11 25/30°C.  
 T ,  
 5°C  
 0, 1, 12, 24, 48, 72 120 -  
 . DNA  
 3- - . A -80°C.  
 C -T 8 FL DNA  
 5' - C - .<sup>14</sup> T  
 (TAP) (CIP) FL DNA -

N . 5 T. L . 287  
 A - sim4<sup>20</sup> U , 1888  
 1943 FL DNA . I DNA , 1943 DNA . O 1888  
 . I , 1360 NCBI  
 (ORF ) DNA . O  
 getorf DNA ( DB) (E < 1 -10; >70% ) . O 1360  
 MPSS 13 EMBOSS .<sup>21</sup> T , 997 DNA  
 N- >70% 1832 DNA ,  
 FL DNA . I CIP TAP FL DNA . O  
 . T  
 PAL2NAL ( (K) , C 0.1.8<sup>22</sup>  
 : 11)<sup>23</sup> , 30% DNA  
 R )<sup>24</sup> ( :// . . /S / FL DNA -  
 RB )<sup>25</sup> ( :// . . / ) . T -  
 .S FOLD ( :// . . )  
 - RNA ' .  
 .. )<sup>26</sup> .<sup>27,28</sup>

3.2. Mapping of the 1888 W1943 FlcDNAs onto cultivated rice *O. sativa* genomic sequences  
 T 1888 FL DNA *O. rufipogon* 1943  
*O. sativa japonica* . N  
 ( 4.0)

**3. Results and Discussion**

3.1. Overall description of W1943 FlcDNA sequences  
 T - DNA *O. rufipogon* (E < 1 -10) G B DB BLAST<sup>5</sup>  
 gon 1943 DNA . O 1888 FL DNA , 1831 (97.0%)  
 .<sup>8</sup> E DNA >80% japonica  
 1 × 10<sup>6</sup> . T DNA (F . 1). T 57 DNA  
 0.5 1.5 . I , . japonica  
 8352 (6432 1920 - 1943 DNA , 395 (21.6%) . A 1831  
 DNA DNA ) 5'- . I japonica . N  
 , 4876 FL DNA 100% . H ,  
 100 1831 DNA , 487  
 Phred >20, . T TGICL <sup>16</sup> 35.8% 1943 DNA DB 100% . T ,  
 4876 DNA . T , DB . I  
 2350 DNA , 454 ( 1896 , - 21.6%),  
 . O 5' (>35.8%) . I  
 3'

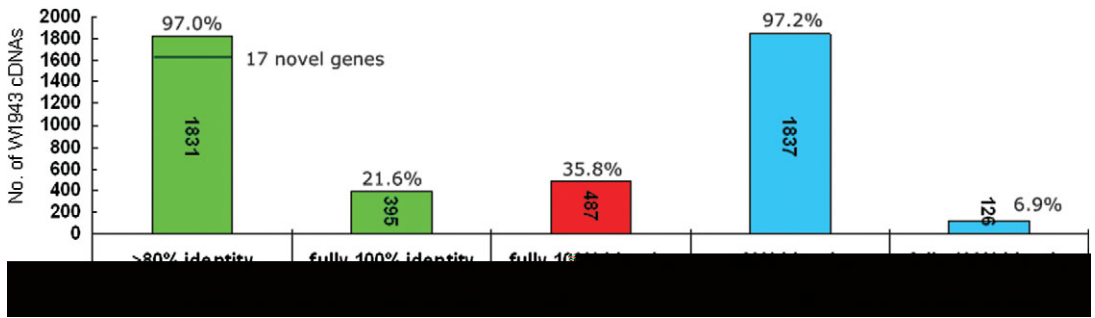


Figure 1. M 1888 FL DNA *Oryza sativa*

288 C C A 1888 F - DNA . 15,  
 O. sativa . indica . 93-11 - 1888 1943 FL DNA I 10 1831 1943  
 BLAST (E < 1 -10). A RNA G B (T 2). A EST  
 1837 (97.2%) 1943 DNA DNA EST RNA DNA  
 . 93-11 >80% (T 3). S  
 (6.9%) (F . 1). O , 126 . A 17 1943 DNA ,  
 . T 1943 - . I ,  
 indica (97.2%) ; . japonica (97.0%) .  
 japonica . indica . H , DNA  
 M .<sup>29</sup> 1943 . japonica . indica DNA ~47 000  
 japonica indica. I . japonica . indica ( :// . . . /).  
 japonica O. rufipogon MPSS ( :// . . . )  
 O. rufipogon .<sup>30</sup> O - / / 17 / / 1943 DNA  
 I 395 1943 FL DNA .<sup>13</sup> T 15 17  
 100% DNA . G  
 EST RNA (T 1). T CU861721' 18  
 15 1943 DNA - ( ) CU406355'  
 EST RNA (T 1). T >100  
 1943. F , I , 57 1943 DNA  
 (CT841942 . japonica . N  
 CU406810) GC-AG . A  
 GT-TG. , 14  
 . indica . 93-11 , 6  
 EST NCBI - , 4  
 Sorghum bicolor, Triticum aestivum,  
 Manihot esculenta Spartina alterniflora EST , 15  
 (F . 2). Gibberella moniliformis, Gibberella

**Table 1.** L 15 *Oryza rufipogon* 1943

A	N	L ( )	C	N	P
CT841942		978	07	6 (1 : GC-AG)	
CU406810		958	06	6 (1 : GT-TG)	D -
CT841893		1011	01	6	D -
CT841874		1369	01	4	
CU405853		1377	05	1	D -
CU405923		639	07	1	IAA
CU406279		648	05	1	
CU406025		839	02	1	
CT841561		740	06	2	
CU406579		468	09	2	
CU406935		1345	01	2	
CU406600		1107	01	2	
CU405570		952	01	2	
CU406091		893	01	3	
CU406134		665	10	3	

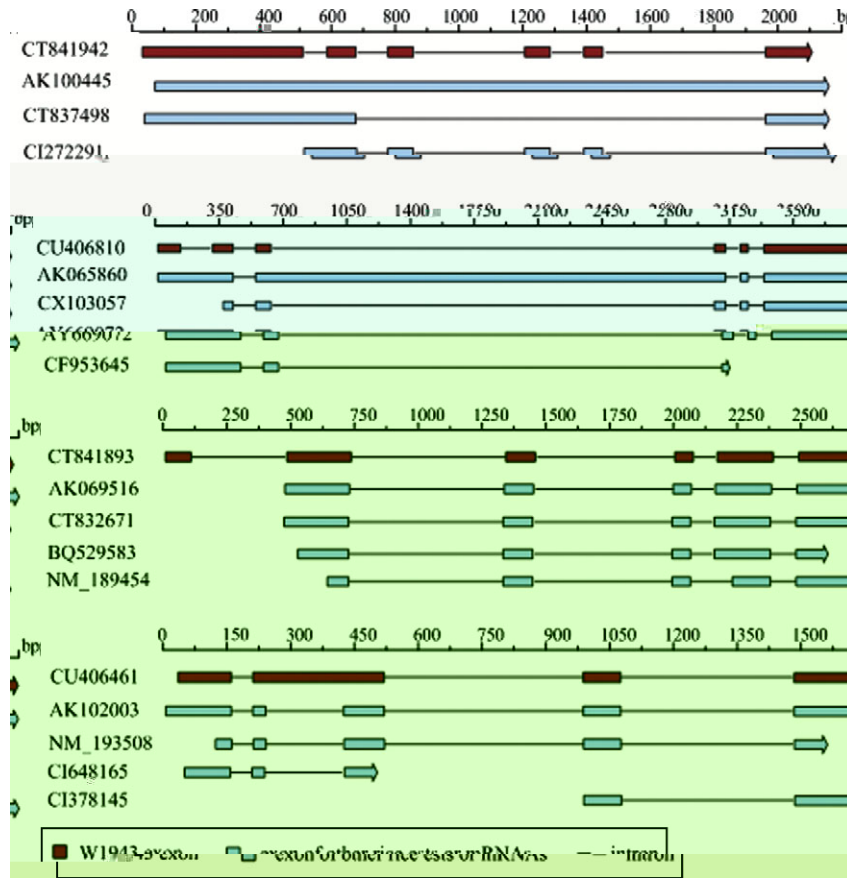


Figure 2. T 17 1943 DNA EST RNA . I

Table 2. L 10 DNA Oryza rufipogon 1943

A	P	L	C	I
N		( )		(%)
CU405785		727	05	99
CU406138		568	02	99
CU406022		543	12	99
CU405757		477	04	100
CU406921		414	02	100
CU406535		389	02	100
CU406832		530	10	92
CU406871		458	01	84
CU861804		383	06	99
CU861721		554	01	100

zeae Magnaporthe grisea, T 4 24 1943 DNA ' 18  
 DNA 18 . S 1943 N 1943

3.3. Comparative analysis with cultivated rice cDNA sequences in public databases

T 1888 1943 DNA DNA  
 japonica . N DNA  
 . indica . G 4 DNA  
 ; :// . . . /RICD).<sup>8</sup>  
 DNA 1888 1943 DNA . F  
 DNA , KOME ( . N  
 4 DNA NCGR (N . G  
 G R , CAS). A , 35 187  
 . japonica FL DNA KOME,  
 . indica FL DNA NCGR.  
 I , DNA  
 N (F . 3).  
 T 1943 DNA , KOME NCGR

A	N	L	( )	P	L	( )	I	(%)	A	L	( )	P
CU405785		727			05			99	CA764081	01		DNA- RNA 3
CU861795		475			09			79	CT858901			U
CU406355		837			12			97	AK107125	12		AP2 ,
CU406396		520			02			99	AK103485	02		H
CT841800		941			11			99	AK121962	11		P ,
CU861688		693			08			99	AK109182	08		H
CT841937		1552			08			98	AK106713	08		U

**Table 4.** L 24 - Oryza sativa . japonica

N	A	N	japonica	93	11	EST	RNA	P
1		CT842002		C	005912	AK241925.1		
2		CT842007		C	008507	CT856206		
3		CU405940		C	001402	AK103326	U	
4		CU406172		C	014596	AK242967.1		
5		CT842006		C	000383	AK111647	GTP-	
6		CU861753		C	000750	AK099287	R -	
7		CU406308		C	000444	AK070131	U	
8		CT841996		C	002576	CT834800	U	
9		CU406568		C	003848	AK064050	B	B
10		CU406582		C	000444	AK107776	U	
11		CU406596		C	001277	AK242711.1	H	
12		CT842008		C	008507	CT856206	U	
13		CU406895		C	003011	CT859459	H	
14		CU861744		C	000750	AK099287	R -	
15		CU405657				CT856885		
16		CT841712				CA766528		
17		CU405768				CT836656	60S	L7A
18		CU405675				CA756235	60S	L17
19		CU406202				NM 001063334	U	
20		CU406924				AC145809		
21		CU405898				CN130755.1 (Sorghum bicolor)	R -	
22		CU406778				BE429292.1 (Triticum turgidum)	H	
23		CU861677				FF534517.1 (Manihot esculenta)	H	
24		CT841912				EH277383.1 (Spartina alterniflora)	U	

DNA . S 1888 1943 DNA . T : 10 , 6  
 1943 MISA ( :// . - - - 5  
 A P / / ) (SSR ) DNA SSR  
 SSR 1 6 DNA , 5'-UTR , ORF

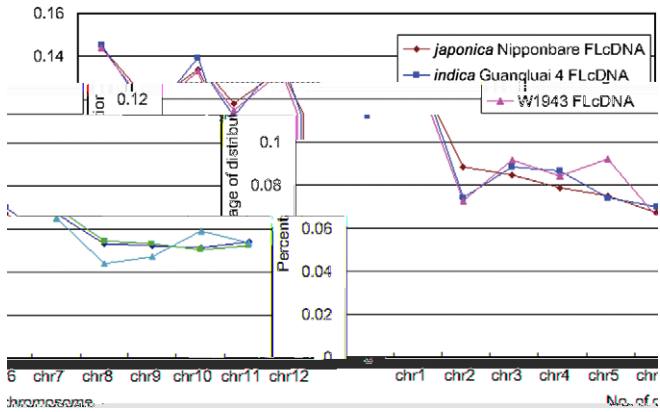


Figure 3. C

DNA ( 1943, KOME, NCGR) . japonica .  
 N . T  
 1943 DNA ,  
 KOME  
 NCGR (KOME, Oryza sativa . japonica N  
 DNA ; NCGR, Oryza sativa . indica G 4 DNA .)  
 3'-UTR , (F . 4).  
 T SSR  
 DNA 5'-UTR, ORF  
 3'-UTR . F , CCG/CGG (54.6%)  
 5'-UTR ORF ,  
 SSR A/T  
 3'-UTR . S , (F . 5B). O 88 , 42 (K )  
 FL DNA . T 1943 KOME; 9  
 CCG/CGG A/T  
 ORF 3'-UTR , 1943 NCGR; 15  
 >50%. H , 5'-UTR , KOME NCGR  
 SSR ≤28%. I , 22  
 SSR - ORF , 76.2% 194  
 UTR . T - SSR - 1943 1943 . N ORF , 59.2%  
 . F ,  
 SSR ; ( 1943/N ) , 0.653 ( 1943/G 4)  
 A/T (7.19%). I , 0.630 (N /G 4),  
 NCGR ORF  
 . indica . japonica. T O. rufipogon 1943 . japonica . indica

1943  
 japonica indica.  
 1943  
 (F . 5). A 823 1943 DNA  
 KOME  
 NCGR (≥95%  
 KOME NCGR).  
 DNA  
 getorf .<sup>21</sup> T  
 194 ORF  
 (F . 5A), 143 ORF  
 1943 KOME, 87 ORF  
 1943  
 NCGR, 64 ORF  
 KOME NCGR. C , 40.9%  
 1943  
 . japonica . N ; 34.1%  
 1943  
 indica . G 4 31.3%  
 . N G 4.  
 T 194 ORF  
 -  
 . T 106 194  
 . S 88  
 (F . 5B). O 88 , 42 (K )  
 1943 KOME; 9  
 NCGR; 15  
 KOME NCGR  
 22  
 , 76.2% 194  
 1943 1943 . N ORF , 59.2%  
 . G 4.  
 29  
 0.302  
 ( 1943/N ) , 0.653 ( 1943/G 4)  
 0.630 (N /G 4),  
 . T ,  
 O. rufipogon 1943 . japonica . indica



Figure 4. T

SSR DNA , 5'-UTR , ORF 3'-UTR

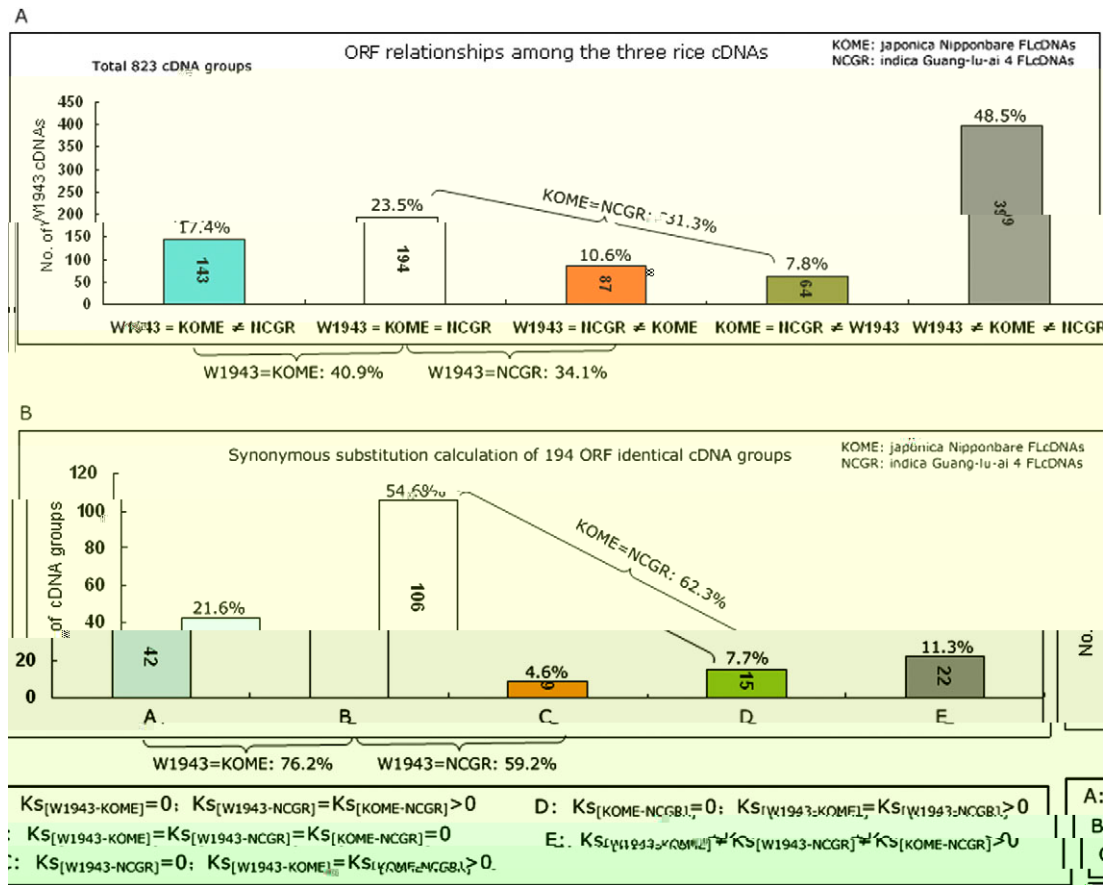


Figure 5. C Oryza sativa DNA (A) T ORF 823 1943, KOME NCGR - DNA (B) T (K) 194 ORF DNA

Table 5. L 4 RNA

A	N	G	( )	P - RNA	( )	H - RNA	RNA	C
CU406292		1416		262 (220 490)		-MIR159		01
CU405943		1511		101 (160 280)		-MIR156		06
CU861819		561		80 (390 470)		- R818		04
CU861752		727		150 (325 475)		- R446		10

3.4. miRNAs identification

A NCBI DB BLAST ,  
432 1888 1943 DNA ,71 - / )  
ORF > 100 , DNA .<sup>13</sup> T -  
361 ,  
- S R : ( ) ,  
RNA ; RB , DNA >100 ; ( )  
-MIR159 , -MIR156 ,  
-MIR818 ; - R446 ,  
(T 5). U FOLD ,  
RNA RNA - ( )  
RNA RNA >10. T , 41  
DNA (T 6). T 16 1943

3.5. Expression analysis by searching against the rice MPSS database

MPSS ( :// . .  
>75%  
>10. T , 41  
(T 6). T 16 1943



**Table 6.** L *Oryza rufipogon* 1943 - ( : )

C	A	L	R	NGS	NCA	NGD	NME	NPO	PFAM A	D	E-
CU406902	44	199	0	101	0	19	0	0	PF07207	L 1	4.8 85
CU405979	36	785	0	894	0	256	9	0			
CT841733	25	112	41	120	0	241	0	0	PF00101	R B CO	2.5 45
CU405975	15	421	1278	0	0	650	223	0			
CT841994	9	140	0	10	0	18	0	0			
CU406521	3	504	6	0	0	0	0	0	PF01070	FMN	2.8 31
CU405996	3	069	0	27	0	28	15	0	PF00430	ATP- B	3.4 28
CU405670	2	653	0	11	5	21	4	23	PF00085	T	7.8 43
CU406006	2	337	0	0	0	0	0	10			
CU406668	2	126	3	17	0	16	0	0			
CT841650	1	997	0	0	0	0	0	0	PF00112	P C1	6 109
CT841731	1	942	0	0	0	12	0	0	PF02507	PSI P F	0
CT841902	1	486	0	24	0	31	0	0			
CU405952	1	253	7	110	5	2	5	0			
CU406199	1	235	0	16	0	0	0	0			
CU406624	1	012	0	60	58	0	3	5	PF05899	DUF861	2.1 37
CU406431	0	189	0	0	0	0	18	17			
CU405706	1	456	15	907	0	183	803	0	PF01439	M 2	2.7 32
CU406330	0	358	4	0	0	0	1	31			
CT841629	2	17	2721	157	36	80	25	86	PF01124	MAPEG	3.1 63
CU406513	1	8	230	0	0	0	0	0	PF01439	M 2	1.6 34
CU406576	0	231	0	11	0	0	0	0			
CU406281	2	9	449	0	0	0	14	0			
CT841966	1	5	520	0	0	0	0	0	PF00188	SCP	5.7 55
CU405942	0	185	0	5	0	0	0	0	PF00967	B	3 84
CU406520	5	1209	0	0	0	0	0	0			
CU406670	0	189	0	0	0	0	0	0	PF00280	P	1.4 20
CU406238	4	1	0	987	33	31	0	0	PF04398	DUF538	4.9 41
CT841875	1	6	0	0	162	3	3	15			
CT841950	1	119	135	76	3079	107	19	0			
CT841815	1	107	135	76	3087	107	19	0			
CU406940	5	9	68	19	31	1393	4	0	PF02065	M	3.5 13
CU406598	5	65	0	606	757	16	965	0	PF00234	T	1.6 31
CU406533	7	0	0	14	30	4662	119	0	PF00234	T	5.5 33
CU406609	0	0	0	0	0	143	0	0			
CU406264	0	0	0	0	0	237	0	0			
CU405759	0	0	0	0	0	779	0	0			
CU406038	1	4	14	0	0	247	0	0			
CU405951	0	25	0	0	0	13	1347	0	PF01439	M 2	6.5 22
CU406698	1	3	0	0	0	0	0	289	PF00481	PP2C	2.4 14
CU406351	1	03	4	36	66	48	42	3228			

NGS, 3 C ; NCA, 35 C ; NGD, 10 G ; NME, 60 C ; NPO,

11 DNA ,1 ,3 CU406902' L 1, ,7 ,1 L 1' L 1 RNA , 2 . S - ,31 A

**Table 7.** L DNA - , - ( : )

C	A	N	NCL	NDL	NSL	PFAM A	D	E-
CU406310		96	2872	3	255	N	N	N
CT841781		96	3089	3	257	N	N	N
CT841558		102	2404	2	365	N	N	N
CU406554		11	568	0	83	N	N	N
CT841576		303	0	3435	68	PF00234	T	4.6 33
CU406485		0	0	1477	0	N	N	N
CU405946		0	113	0	591	PF00257	D	2.2 54
NCL, 14			4°C		24 ; NDL, 14		5 ; NSL,	
14			250 M N C		24 .			

CT841733' R B CO ' ( -1,5- ) . A R B CO ' T ' R B CO 32 I 1943 DNA (T 7). O DNA , - . I PFAM D ' T 33

3.6. Conclusions I 1888 FL DNA O. rufipogon G . 1943. A 17 DNA 41 . T O. rufipogon 1943 O. sativa . japonica . indica - . I 1943 (26°14'N, 116°36'E) J P C 34 I O. rufipogon 35 B O. sativa . japonica indica 1943 . T

**Acknowledgements:** P G C (T , J ) 1943.

**Funding** T M S T C ( C R F G P , 2005CB120805 2006AA10A102), A S (038019315 KSC02- N-024) S M C X S T .

**References**

1. P , . , S , G. T , S. D. 1992, Oryza RFLP , Theor. Appl. Genet., **83**, 565 581.
2. L , J. P, C , . C, H , K. H., C , T. . S , B. A. 2006, P A , Oryza rufipogon, , Oryza sativa, Proc. Natl. Acad. Sci. USA, **103**, 9578 9583.
3. , O , S, F , , S , , , O S , C. 2006, I D (O. rufipogon G .), Plant Mol. Biol., **62**, 247 259.
4. T , F, , , B, . 2006, F (Oryza rufipogon G .), Theor. Appl. Genet., **113**, 619 629.
5. I R G S P , 2005, T - , Nature, **436**, 793 800.
6. , J, H , S, , J, . 2002, A Oryza sativa L . indica, Science, **296**, 92 100.
7. T R F -L DNA C , 2003, C , 28,000 DNA japonica , Science, **301**, 376 379.
8. L , O, L , T, , S, . 2007, A 10,096 DNA Oryza sativa indica japonica , Plant Mol. Biol., **65**, 403 415.
9. M N , K. L., B , R, M , D, B , C. R., L , J. E. L , H. 2006, S

10. S , K., D , K., N , T., . 2007, G DNA , PLoS ONE, **2**, 1235.
11. C , S. K., O , S. H., J , J. U., . 2004, C 5,211 EST (Oryza minuta), Plant Cell Rep., **22**, 839 847.
12. M , H., S , O , H. I. 1992, E , Oxford Surv. Evol. Biol., **8**, 135 184.
13. N , M., N , K., , K., T , S. S., S , J. . M , B. C. 2006, P MPSS RNA RNA, Nucleic Acids Res., **34**, 731 735.
14. C , P., K , C., K , A., . 1996, H DNA CAP , Genomics, **37**, 327 336.
15. E , B. G , P. 1998, B . II. E , Genome Res., **8**, 186 194.
16. P , G., H , O. L , F., . 2003, TIGR I (TGICL): EST , Bioinformatics, **19**, 651 652.
17. A , S. F., M , T. L., S , A. A., . 1997, G BLAST PSI-BLAST: P , Nucleic Acids Res., **25**, 3389 3402.
18. A , R., A , T. K., B , A., . 2001, T I P , Nucleic Acids Res., **29**, 37 41.
19. B , A., C , L., D , R., . 2004, T P , N A R ., **32**, 138 141.
20. F , L., H , G., , R , G. M. , M. 1998, A DNA DNA , Genome Res., **8**, 967 974.
21. R , P., L , I. B , A. 2000, EMBOSS: E , Trends Genet., **16**, 276 277.
22. T , J. D., G , T. J., P , F., J , F. H , D. G. 1997, T C O , Nucleic Acids Res., **24**, 4876 4882.
23. S , M., T , D. B , P. 2006, PAL2NAL: , Nucleic Acids Res., **34**, 609 612.
24. G -J , S., M , S., M , M., K , A., E , S. R. B , A. 2005, R : RNA , Nucleic Acids Res., **33**, D121 D124.
25. G -J , S., S , H. K., D , S. E , A. J. 2008, RB : RNA , Nucleic Acids Res., **36**, D154 D158.
26. , M. 2003, M , Nucleic Acids Res., **31**, 3406 3415.
27. T , K., I , T., , R., . 2005, D : , Genome Res., **17**, 1005 1014.
28. C , P., S , A., L , B., . 2006, G - , Nat Genet., **38**, 626 635.
29. M , L., O , R., M , H., K , A. M , . 2006, G - (Oryza sativa L.) - (Oryza rufipogon G .), DNA Res., **13**, 43 51.
30. C , C., M , R., T , S., F , , O , H. O , E. 2003, P : SINE , Mol. Biol. Evol., **20**, 67 75.
31. R , C. D , R. 1993, C - , Plant Mol. Biol., **22**, 165 170.
32. T , N. E., C , . G., T , G. J., H , C. M., F , R. T. S , D. M. 1986, T -1,5- , Nucleic Acids Res., **14**, 3325 3342.
33. C , T. J., K , A. A. C , P. M. 1989, A DNA- ( ) , Plant Mol. Biol., **13**, 95 108.
34. G , L. ., H , D. . G , S. 2000, A - Oryza rufipogon G . C , Theor. Appl. Genet., **101**, 494 502.
35. , . S., , L. H., L , . , O. K. 1996, G RFLP ( C ), Agric Biotechnol., **4**, 111 117.