

水稻粒宽基因 *GW5* (*qSW5*) 的生物信息学分析

第七组:

李文清 刘尧 宁东峰 张
伶燕 张文学 刘家范



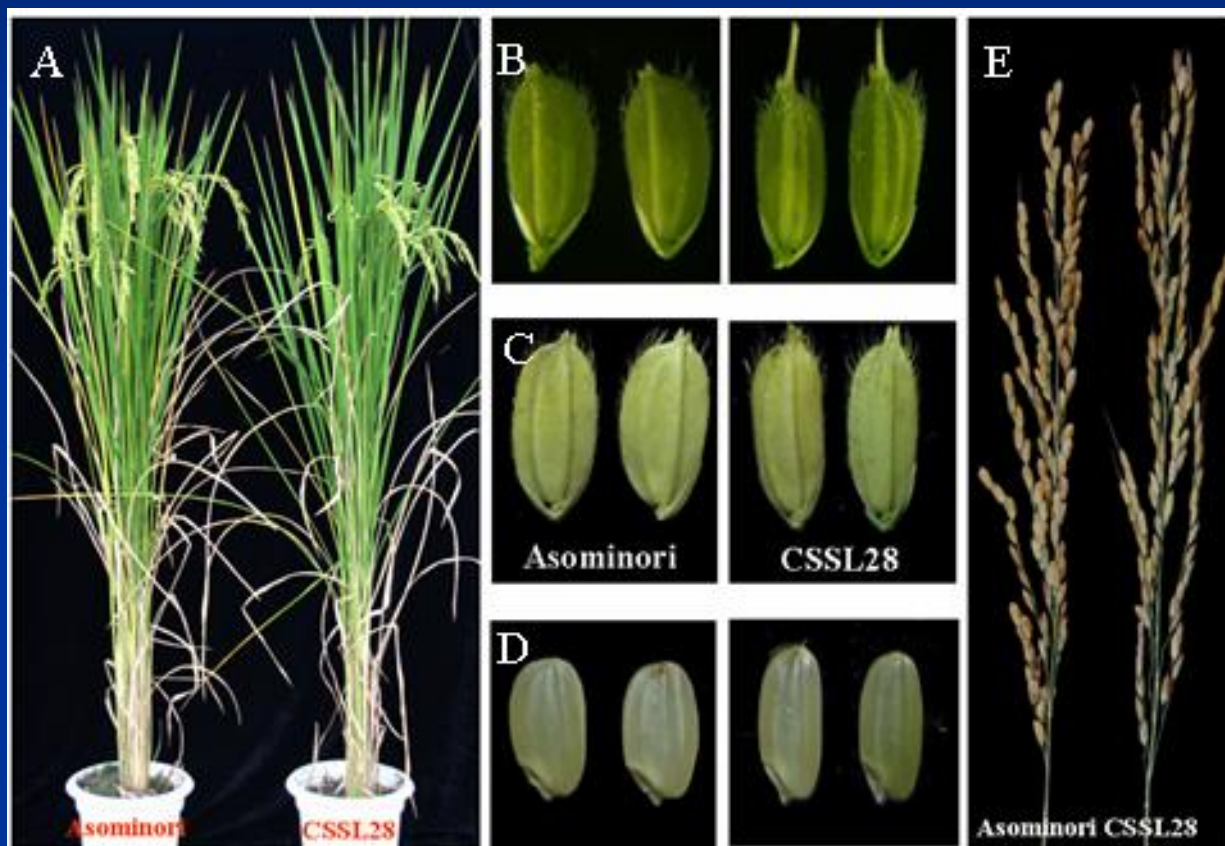
第一部分：研究背景

谷粒重量是构成水稻产量的三要素之一，它是籽粒库容（长、宽、厚）及籽粒充实度的综合指标，一般以千粒重表示。在水稻育种中，提高籽粒重量是一条有效的增产途径。

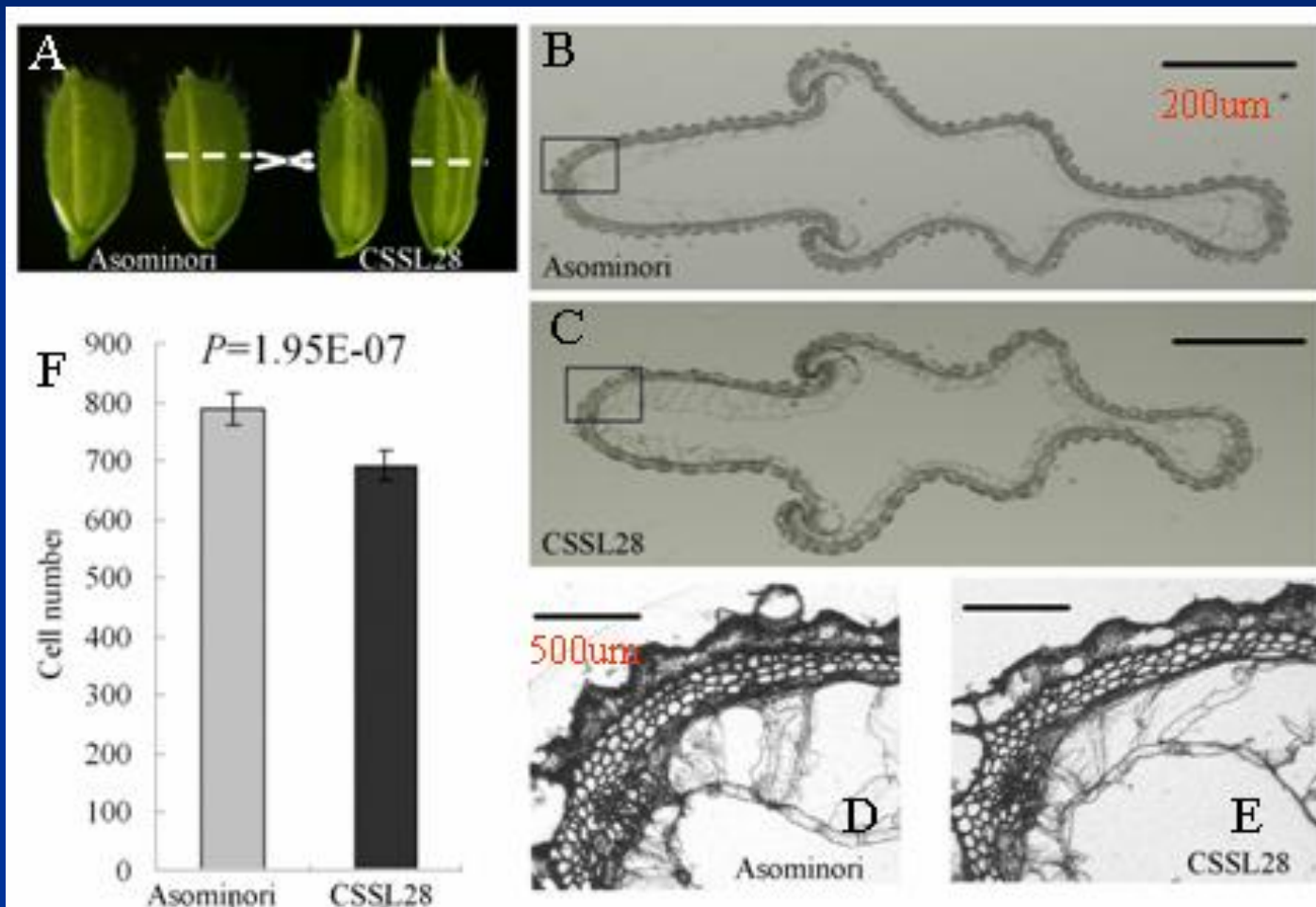
万建民实验室通过Asominori与CSSL28构建F₂分离群体，图位克隆方法分离该QTL，利用染色体片段置换系材料精细定位了多环境、多遗传背景稳定表达并控制粒宽、粒重基因GW5

形态学研究表明：GW5能增加颖壳外围薄壁细胞数目，进而提高库容及同化产物运输能力，导致水稻粒宽差异。对于高产与品质育种具有重要的指导意义

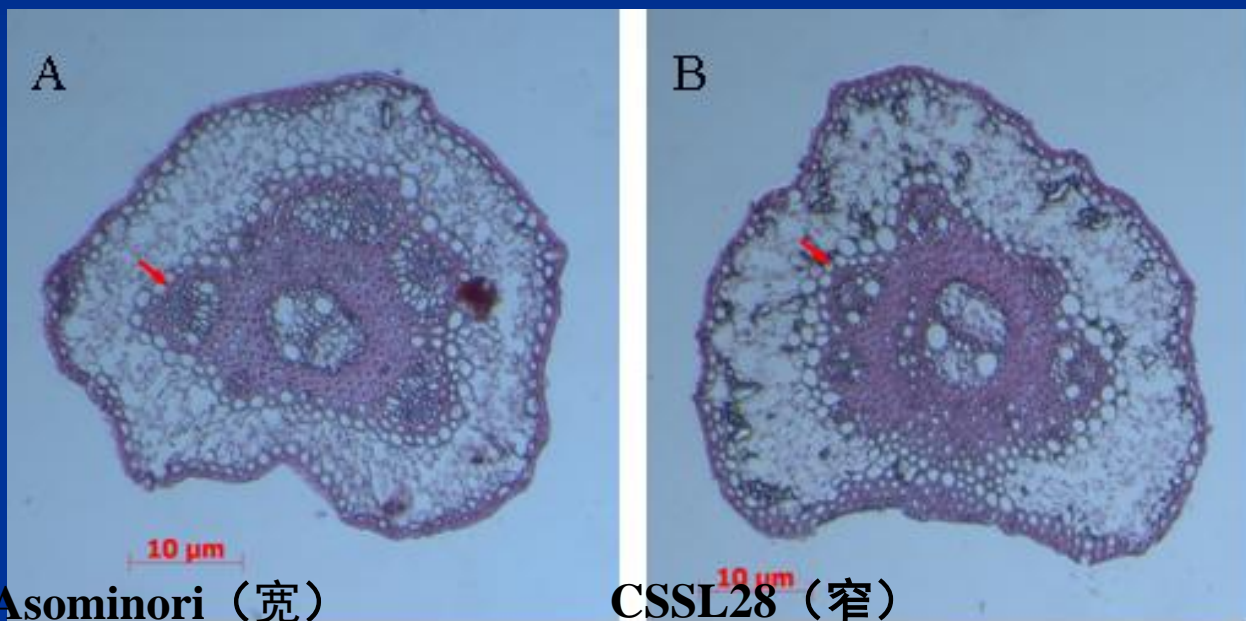
置换系与背景亲本间粒宽差异显著



颖壳外围薄壁细胞数目差异分析



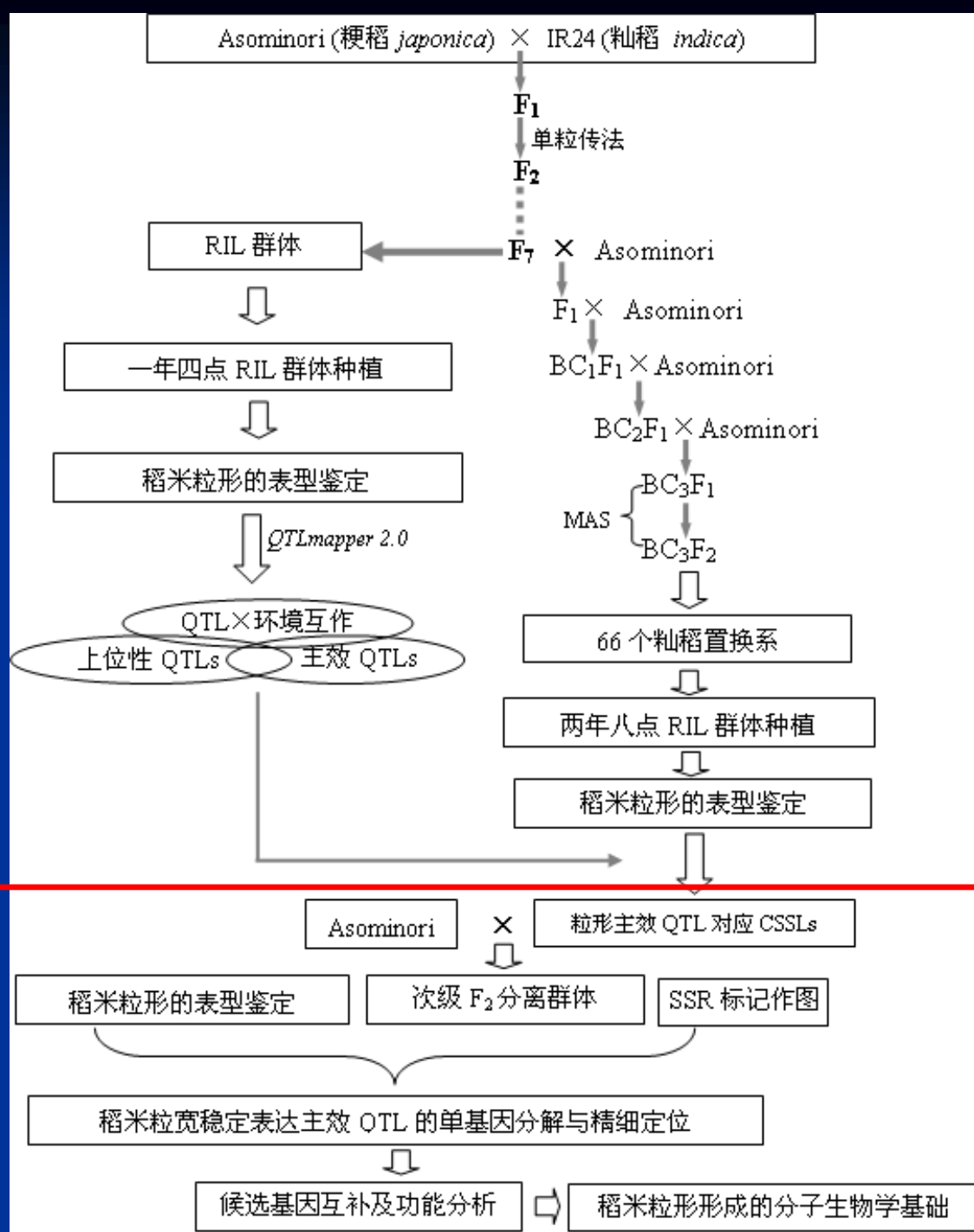
颖壳果柄的维管束横切比较



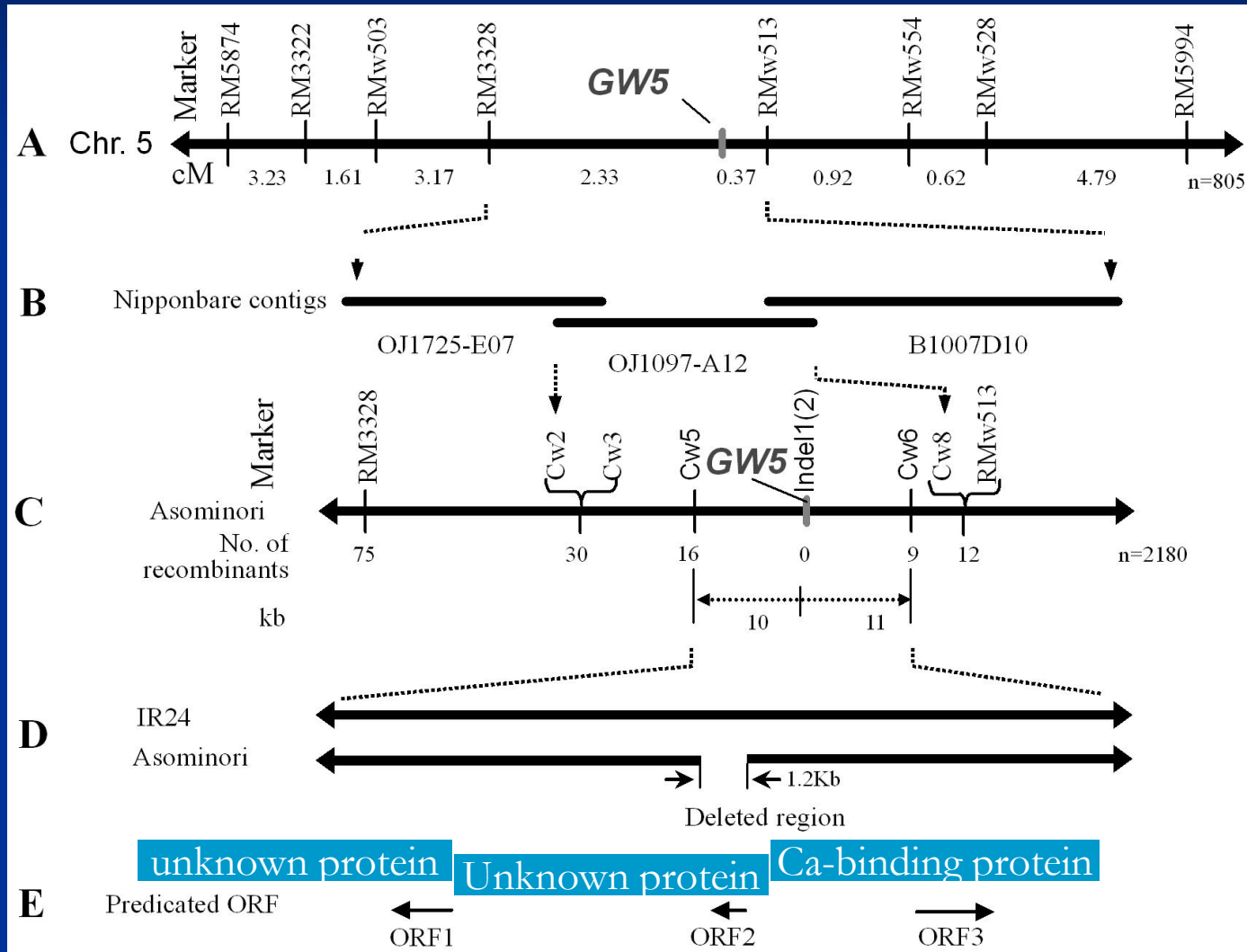


第二部分：已有实验数据

研究技术路线

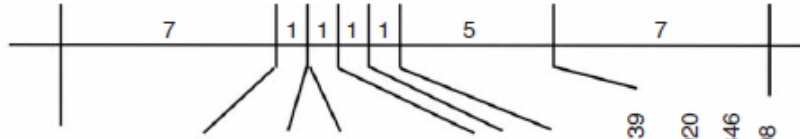


粒宽基因GW5的精细定位



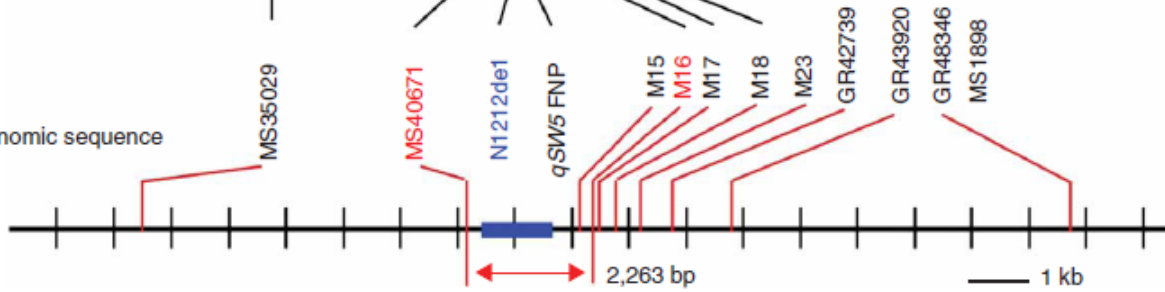
a

Genetic map
(*n* = 4,501)



b

Kasalath genomic sequence



c

BK fragment



EN fragment



XK fragment



EE fragment



d

Predicted ORF1



Predicted ORF2



Predicted ORF3

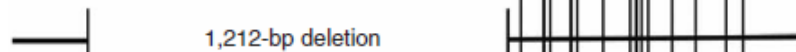


Fragment for RNAi

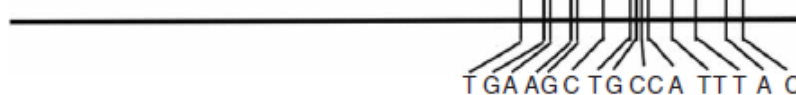


e

Nipponbare



Kasalath



2,263-bp candidate region

CAGGATCA TTTGGCTT
 TGAAGCTGCCA TTTA C

Mapping and cloning of *qSW5*



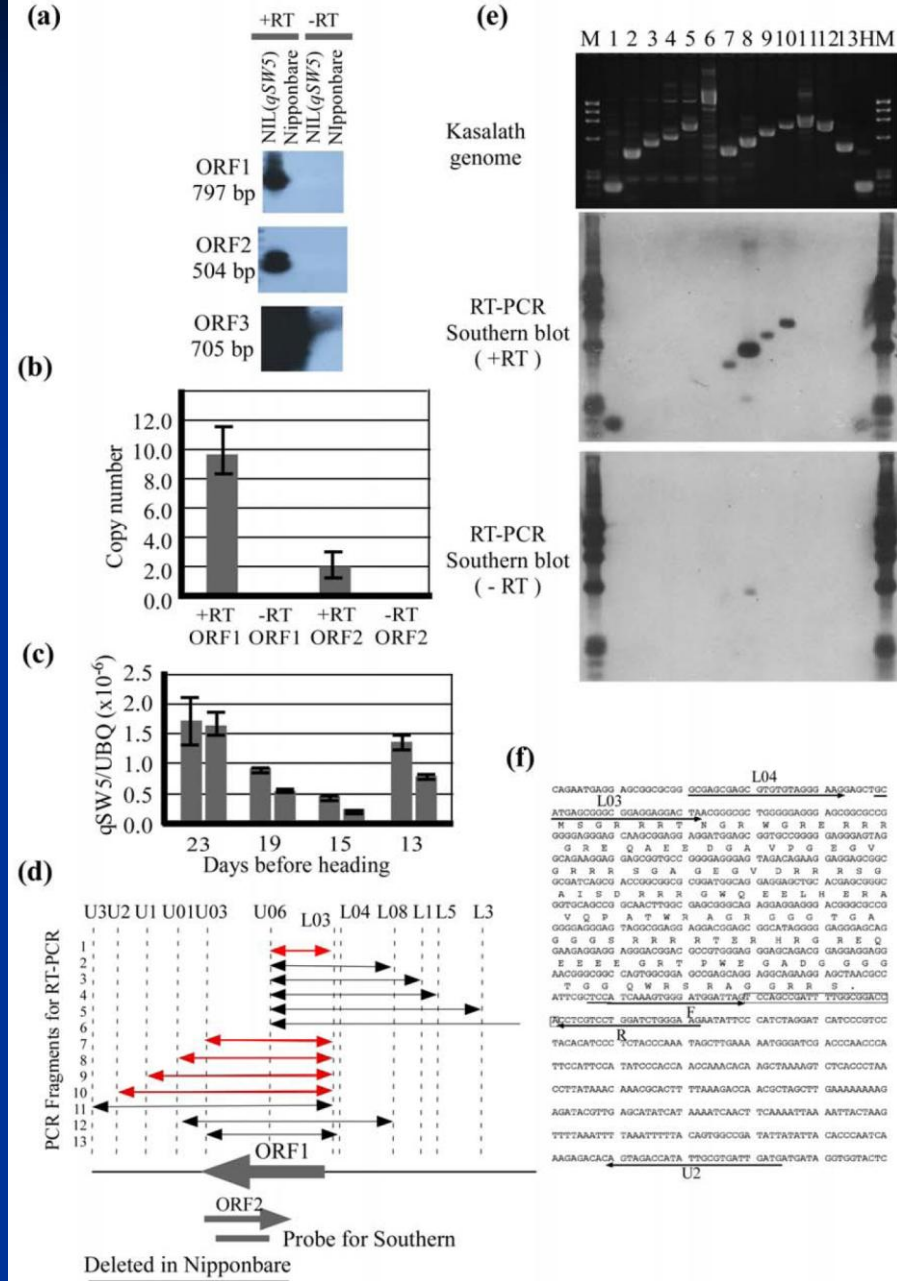
Shomura, *et al*,
Nature genetics,
 2008,40: 1023-1028

Mapping and cloning of qSW5

- (a) Genetic map for mapping of qSW5, made with 4,501 segregating plants. Numbers indicate recombination events between DNA markers indicated in the figure.
- (b) Kasalath physical genomic map corresponding to the genetic map in a, with DNA marker positions. qSW5 FNP was delimited to a 2,263-bp region.
- (c) Four Kasalath genome fragments used for the complementation test.
- (d) Predicted ORFs in the qSW5 candidate regions. Of the three ORFs, two contained no intron but were long ORFs in the opposite orientation in the same region. ORF1 is the qSW5 gene product candidate based on the expression analysis (Supplementary Fig. 3).
- (e) Polymorphisms between Kasalath and Nipponbare in the candidate region. A major difference was a 1,212-bp deletion.

qSW5 expression analysis and identification of *qSW5* transcripts

Supplementary Fig. 3



- (a) ORF1 and 2 were expressed only in NIL(*qSW5*) samples positive for reverse transcriptase.
- (b) Confirmation of direction of *qSW5* transcription. ORF1 and ORF2 were predicted in opposite directions from the same genomic region (c) ORF1 mRNA expression by Taq-man quantitative RT-PCR.
- (d) Schematic positions of primers for RT-PCR.
- (e) Results of RT-PCR for ORF1.
 Top: Genome control to confirm amplification under the same PCR conditions.
 Middle: RT-PCR followed by Southern blot analysis.
 Bottom: RT-PCR followed by Southern blot analysis.
- (f) cDNA sequence amplified by RT-PCR. cDNA was amplified by using primers L03 and U2, not primers L04 and U2. Predicted amino acid sequences of ORF1

小结

通过已发表实验室数据，对*GW5(qSW5)*我们可以得到如下结论：

- 1、功能互补实验中，只有11.2K一条片段能够恢复功能，可确定该控制粒宽的基因存在于此11.2K片段中，但已发表的预测基因*GW5(qSW5)*均不完整。
- 2、转录分析、RNAi实验及籼粳间序列比较可确定该控制粒宽的基因包含籼粳间缺失的1.2K序列。
- 3、该控制粒宽的基因转录方向与*GW5(qSW5)*一致。
- 4、*qSW5*预测区域上游可能存在长的内含子
- 5、*qSW5*预测区域上游可能存在一个或多个与该基因共同作用的基因

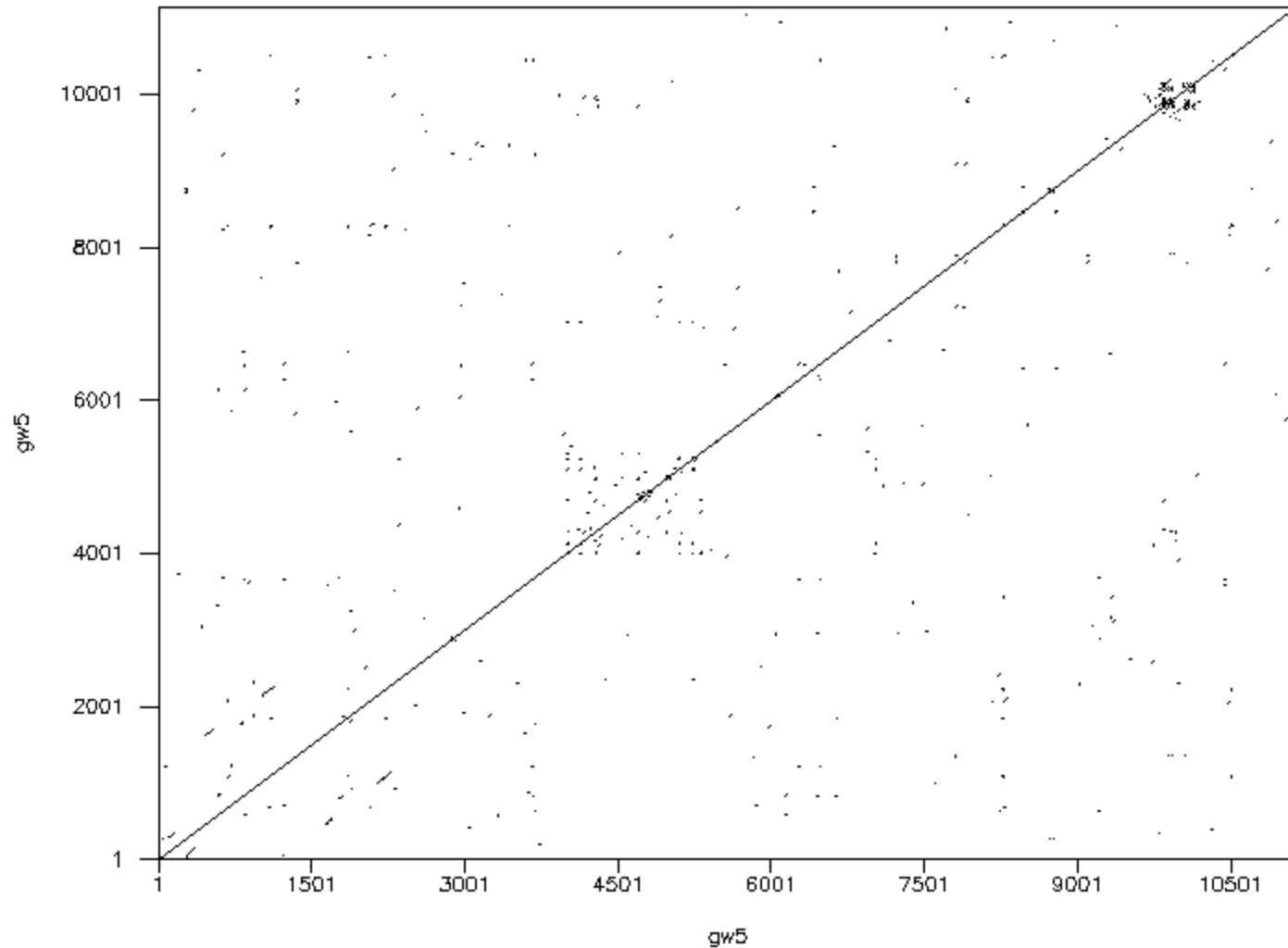


第三部分：生物信息分析

Dottup 分析重复序列

Dottup: fasta::490661:gw5 vs fasta::490662:gw5

Wed 23 Nov 2011 13:15:13



BLST 分析同源序列

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
AB433345.1	Oryza sativa Indica Group qSW5 gene, complete cds	5591	5591	100%	0.0	100%	
AC093954.2	Oryza sativa Japonica Group chromosome 5 clone OJ1097_A12, comp	3692	3692	72%	0.0	96%	
HQ658632.1	Oryza sativa Indica Group isolate Habataki qSW5 (qSW5) gene, comp	1631	1631	30%	0.0	98%	
HQ658631.1	Oryza sativa Indica Group isolate Zhenshan 97 qSW5 (qSW5) gene, c	1631	1631	30%	0.0	98%	
HQ658633.1	Oryza sativa Japonica Group isolate 02428 qSW5 (qSW5) gene, comp	1622	1622	30%	0.0	98%	
DQ991205.1	Oryza sativa (indica cultivar-group) GW-5 (GW-5) gene, complete c	1519	1519	28%	0.0	98%	
AP005765.3	Oryza sativa Japonica Group genomic DNA, chromosome 9, BAC clone	614	736	15%	1e-171	96%	
AP005159.3	Oryza sativa Japonica Group genomic DNA, chromosome 8, BAC clone	394	449	16%	2e-105	80%	
AF488413.1	Oryza sativa chromosome 6 BAC 134P10, complete sequence	280	496	24%	3e-71	76%	
AP002542.2	Oryza sativa Japonica Group genomic DNA, chromosome 6, PAC clone	280	496	24%	3e-71	76%	
AP000559.1	Oryza sativa Japonica Group genomic DNA, chromosome 6, clone:P0	280	496	24%	3e-71	76%	
AP004553.2	Oryza sativa Japonica Group genomic DNA, chromosome 8, BAC clone	269	269	17%	6e-68	72%	
AP003540.2	Oryza sativa Japonica Group genomic DNA, chromosome 6, PAC clone	266	266	17%	8e-67	72%	
AP005387.3	Oryza sativa Japonica Group genomic DNA, chromosome 6, BAC clone	251	251	17%	2e-62	71%	
AP004583.3	Oryza sativa Japonica Group genomic DNA, chromosome 6, BAC clone	223	223	7%	8e-54	82%	
AP003743.4	Oryza sativa Japonica Group genomic DNA, chromosome 7, BAC clone	219	219	7%	1e-52	80%	
AP003316.4	Oryza sativa Japonica Group genomic DNA, chromosome 1, PAC clone	214	214	6%	4e-51	84%	
AP003407.3	Oryza sativa Japonica Group genomic DNA, chromosome 1, BAC clone	214	214	6%	4e-51	84%	
AP004613.3	Oryza sativa Japonica Group genomic DNA, chromosome 1, BAC clone	203	203	7%	8e-48	80%	
AP003284.3	Oryza sativa Japonica Group genomic DNA, chromosome 1, PAC clone	190	190	9%	5e-44	76%	
AP003222.3	Oryza sativa Japonica Group genomic DNA, chromosome 1, PAC clone	190	190	9%	5e-44	76%	
AP004791.3	Oryza sativa Japonica Group genomic DNA, chromosome 6, PAC clone	187	187	8%	6e-43	77%	
CR855138.1	Oryza sativa genomic DNA, chromosome 4, BAC clone: OSIGBa0118P	179	179	7%	9e-41	78%	
AP005719.3	Oryza sativa Japonica Group genomic DNA, chromosome 7, BAC clone	178	178	16%	3e-40	70%	
AP005160.3	Oryza sativa Japonica Group genomic DNA, chromosome 7, BAC clone	178	178	16%	3e-40	70%	
AP003282.2	Oryza sativa Japonica Group genomic DNA, chromosome 1, PAC clone	178	178	9%	3e-40	74%	
AP003767.3	Oryza sativa Japonica Group genomic DNA, chromosome 6, PAC clone	172	172	9%	1e-38	74%	
AP003487.2	Oryza sativa Japonica Group genomic DNA, chromosome 6, BAC clone	172	172	9%	1e-38	74%	
AL606638.2	Oryza sativa genomic DNA, chromosome 4, BAC clone: OSJNBa0041A	172	172	7%	1e-38	78%	
AP003344.2	Oryza sativa Japonica Group genomic DNA, chromosome 1, PAC clone	172	172	8%	1e-38	76%	

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
ABJ90467.1	GW-5 [Oryza sativa Indica Group] >gb EAY96821.1 hypothetical protein Osl_18745 [Oryza sativa Indica Group]	142	142	27%	5e-40	92%
BAD61769.1	hypothetical protein [Oryza sativa Japonica Group] >gb EEE66733.1 hypothetical protein OsJ_23424 [Oryza sativa Japonica Group]	79.0	79.0	18%	4e-14	70%
BAG54844.1	qSW5 [Oryza sativa Indica Group]	73.2	73.2	17%	2e-13	100%
AEI87384.1	qSW5 [Oryza sativa Indica Group] >gb AEI87385.1 qSW5 [Oryza sativa Indica Group] >gb AEI87386.1 qSW5 [Oryza sativa Japonica Group]	71.6	71.6	17%	6e-13	96%
BAD68530.1	hypothetical protein [Oryza sativa Japonica Group]	70.1	70.1	21%	5e-11	58%
BAD69021.1	hypothetical protein [Oryza sativa Japonica Group]	58.2	58.2	15%	3e-07	68%
BAD53238.1	hypothetical protein [Oryza sativa Japonica Group]	57.4	57.4	15%	4e-07	68%
EEC83398.1	hypothetical protein Osl_28836 [Oryza sativa Indica Group]	57.4	57.4	20%	6e-07	55%
EEE61537.1	hypothetical protein OsJ_15855 [Oryza sativa Japonica Group]	57.0	57.0	15%	8e-07	70%
BAA84811.1	hypothetical protein [Oryza sativa Japonica Group] >db BAB19373.1 hypothetical protein [Oryza sativa Japonica Group]	55.1	55.1	19%	9e-07	53%
EEC77834.1	hypothetical protein Osl_17057 [Oryza sativa Indica Group]	57.0	57.0	15%	1e-06	70%
EEE68509.1	hypothetical protein OsJ_26934 [Oryza sativa Japonica Group]	55.1	55.1	20%	4e-06	54%
BAD68252.1	hypothetical protein [Oryza sativa Japonica Group]	52.0	52.0	15%	1e-05	67%
EAY92089.1	hypothetical protein Osl_13795 [Oryza sativa Indica Group]	50.8	50.8	9%	6e-05	77%
BAD54413.1	hypothetical protein [Oryza sativa Japonica Group]	50.8	50.8	9%	8e-05	77%
BAD01716.1	hypothetical protein [Oryza sativa Japonica Group]	51.2	51.2	10%	8e-05	71%
EEC70479.1	hypothetical protein Osl_01542 [Oryza sativa Indica Group]	48.1	48.1	17%	9e-04	46%
BAD54639.1	hypothetical protein [Oryza sativa Japonica Group] >db BAD69394.1 hypothetical protein [Oryza sativa Japonica Group]	47.4	47.4	21%	0.002	43%
BAC83221.1	hypothetical protein [Oryza sativa Japonica Group]	42.4	42.4	13%	0.005	48%
BAD17768.1	hypothetical protein [Oryza sativa Japonica Group] >db BAD17776.1 hypothetical protein [Oryza sativa Japonica Group]	43.1	43.1	16%	0.015	42%
ABA95657.1	hypothetical protein LOC_Os12g02610 [Oryza sativa Japonica Group] >gb EAZ19459.1 hypothetical protein OsJ_35022 [Oryza sativa Japonica Group]	43.5	43.5	17%	0.020	40%
EAY74838.1	hypothetical protein Osl_02730 [Oryza sativa Indica Group]	42.0	42.0	19%	0.081	42%
EEE60451.1	hypothetical protein OsJ_13695 [Oryza sativa Japonica Group]	42.0	42.0	24%	0.096	38%
BAD46361.1	hypothetical protein [Oryza sativa Japonica Group]	41.6	41.6	16%	0.10	40%
EEE70083.1	hypothetical protein OsJ_30078 [Oryza sativa Japonica Group]	41.6	41.6	16%	0.10	40%
AAK92659.1	Hypothetical protein [Oryza sativa Japonica Group]	41.6	41.6	22%	0.15	41%
EEE50839.1	hypothetical protein OsJ_31263 [Oryza sativa Japonica Group]	39.3	39.3	16%	0.18	42%
EAY99528.1	hypothetical protein Osl_21497 [Oryza sativa Indica Group]	39.7	39.7	13%	0.50	46%
EEC79182.1	hypothetical protein Osl_19877 [Oryza sativa Indica Group]	38.9	38.9	19%	0.51	41%
EEE55194.1	hypothetical protein OsJ_03038 [Oryza sativa Japonica Group]	39.3	39.3	16%	0.73	40%
EAZ43765.1	hypothetical protein OsJ_28384 [Oryza sativa Japonica Group]	38.5	38.5	22%	0.89	35%
EEE63646.1	hypothetical protein OsJ_18463 [Oryza sativa Japonica Group]	38.1	38.1	19%	0.93	41%
EEE68526.1	hypothetical protein OsJ_26968 [Oryza sativa Japonica Group]	36.2	57.4	14%	1.4	75%
ZP_01997414.1	hypothetical protein BGS_0560 [Beggiatoa sp. SS] >gb EDN72584.1 hypothetical protein BGS_0560 [Beggiatoa sp. SS]	35.4	35.4	12%	3.0	40%
EFR24063.1	hypothetical protein AND_11639 [Anopheles darlingi]	37.0	37.0	27%	5.4	28%
YP_004155225.1	serine/threonine protein kinase [Variovorax paradoxus EPS] >gb ADU37114.1 serine/threonine protein kinase [Variovorax paradoxus EPS]	37.0	37.0	9%	5.7	58%
ZP_03927280.1	polyphosphate kinase [Actinomyces urogenitalis DSM 15434] >gb EEH65853.1 polyphosphate kinase [Actinomyces urogenitalis DSM 15434]	36.6	36.6	24%	7.7	35%
AEY84133.1	DNA-directed RNA polymerase II subunit RPB1 [Actinoplanes sp. SE50/110]	35.4	35.4	9%	9.2	52%

BLSTx

分析同源序列

Gn.Ex	Type	S	.Begin	...End	.Len	Fr	Ph	I/Ac	Do/T	CodRg	P....	Tscr..
1.01	PlyA	+	613	618	6							2.27
2.05	PlyA	-	744	739	6							2.27
2.04	Term	-	4852	3806	1047	1	0	67	45	2796	0.991	273.88
2.03	Intr	-	5423	4939	485	0	2	76	107	1447	0.867	143.19
2.02	Intr	-	7052	6957	96	0	0	17	45	150	0.650	6.30
2.01	Init	-	7180	7144	37	1	1	126	-1	-19	0.708	-4.16
2.00	Prom	-	7396	7357	40							0.72
3.00	Prom	+	7617	7656	40							-0.78
3.01	Init	+	7793	7986	194	1	2	71	5	155	0.578	7.34
3.02	Intr	+	9633	9737	105	0	0	8	43	152	0.172	4.80
3.03	Term	+	9796	10198	403	1	1	-98	35	639	0.909	39.31
3.04	PlyA	+	10444	10449	6							0.77
4.01	PlyA	-	10819	10814	6							2.27

基因预测

根据实验结论，预测的基因中，仅有此**预测基因3**与根据实验所做猜测想符合，在*qSW5*预测区域上游存在一长的内含子（1.8K）

另外**预测基因2**可能为上游共同作用的基因。

FGENESH 2.6 Prediction of potential genes in Monocot genomic DNA

Seq name: test sequence

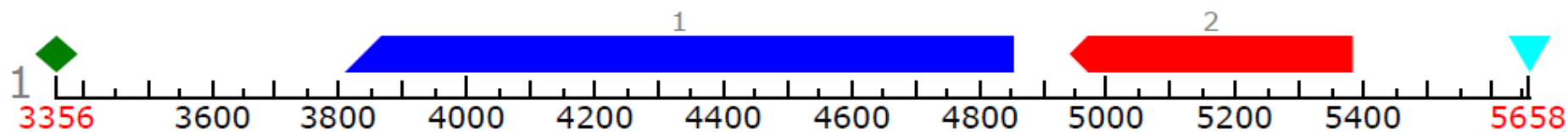
Length of sequence: 11150

Number of predicted genes 2: in +chain 0, in -chain 2.

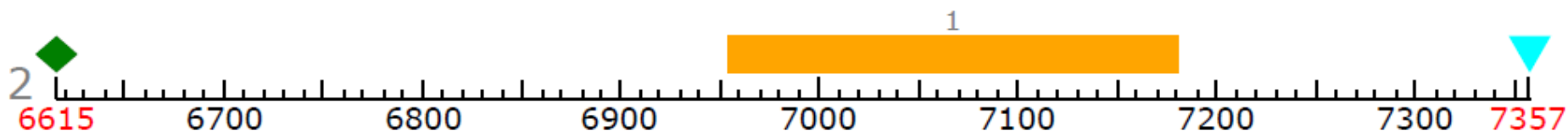
Number of predicted exons 3: in +chain 0, in -chain 3.

Positions of predicted genes and exons: Variant 1 from 1, Score:102.662012

▶ CDSf
 ■ CDSi
 ▴ CDSl
 ■ CDSo
 ◆ PoIA
 ▼ TSS

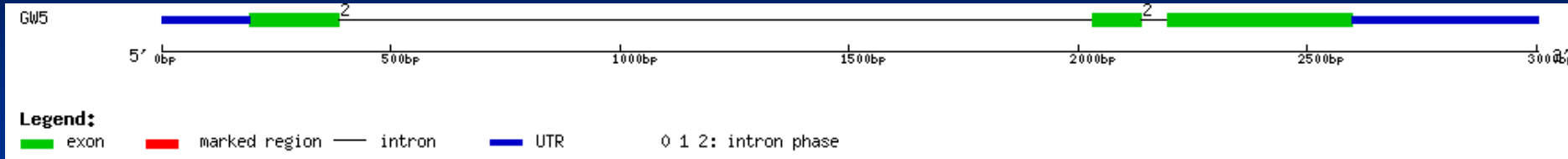


1	-	PoIA	3356		0.44					
1	-	1 CDSl	3806	-	4852	71.58	3806	-	4852	1047
1	-	2 CDSf	4939	-	5382	39.64	4939	-	5382	444
1	-	TSS	5658		-8.38					



2	-	PoIA	6615		0.44					
2	-	1 CDSo	6953	-	7180	7.00	6953	-	7180	228
2	-	TSS	7357		-5.38					

预测基因3及其氨基酸序列




>/tmp/11_14_11-

08:26:01.fasta | GENSCAN_predicted_peptide_3 | 233_aa

MHRGRRGGDKLWPWTKDELPHRTGAAERRRGGGDRGEG
KSCGQRRSRHTTASEAGRGRHLDTASDEEGAANGGRWKQE
RQTTGDRRRGWKERCASEQAQASCMSGRRRTNGRWGRERR
RGREQAEEDGAVPGEGVGRRRSGAGEGVDRRRSGAISDRRR
GWQEELHERAVQPATWRAGRGGGTGAGGGSRRRRRTERHR
GREQEEEEGRTPWEGADGGGGGTGGQWRSRAGGRRS

预测基因3氨基酸序列BLST

Graphical overview

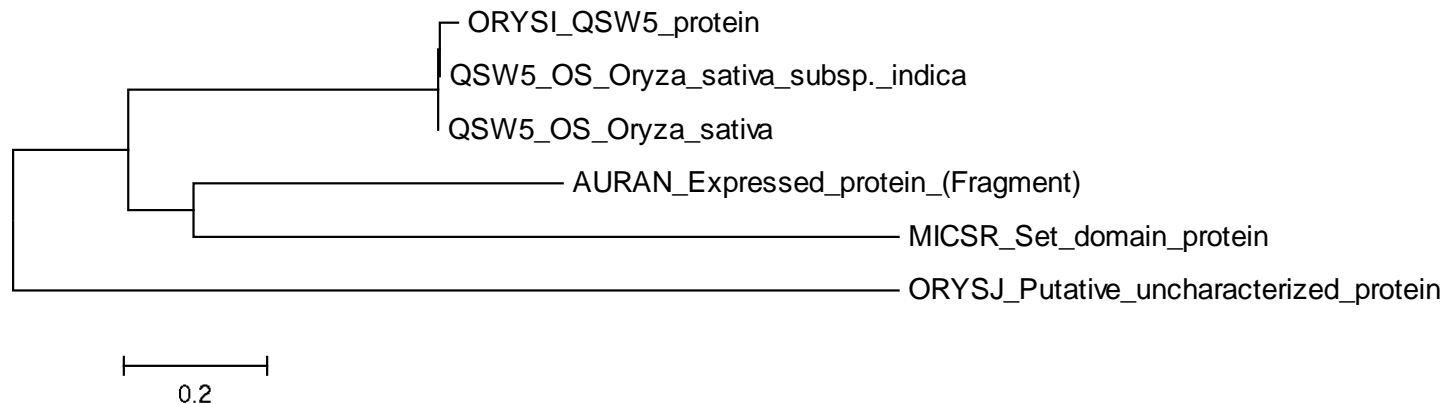
Color code for identity 0
 .100% = 

Accession	Entry name	0Query hit233	0Match hit (sqrt scale)1106	Name (Organism)
<input type="checkbox"/>	Query 20111224612AKFZ8VJ			
<input checked="" type="checkbox"/>	B3IXB4 B3IXB4_ORYSI			QSW5 protein (Oryza sativa subsp. indica)
<input checked="" type="checkbox"/>	F8SPD1 F8SPD1_ORYSJ			QSW5 (Oryza sativa subsp. japonica)
<input checked="" type="checkbox"/>	F8SPC9 F8SPC9_ORYSI			QSW5 (Oryza sativa subsp. indica)
<input checked="" type="checkbox"/>	Q75KY6 Q75KY6_ORYSJ			Putative uncharacterized protein... (Oryza sativa subsp. japonica)
<input checked="" type="checkbox"/>	F0Y057 F0Y057_AURAN			Expressed protein (Aureococcus anophagefferens)
<input checked="" type="checkbox"/>	C1EG84 C1EG84_MICSR			Set domain protein (Micromonas sp. (strain RCC299 / NOUM17))

Detailed BLAST results [Customize](#)

Show hits from [complete proteomes only](#).

Alignments	Entry	Entry name	Status	Protein names	Organism	Length	Identity	Score	E-value	Gene names
<input checked="" type="checkbox"/>	B3IXB4	B3IXB4_ORYSI	★	QSW5 protein	Oryza sativa subsp. indica (Rice)	131	71.0%	178	3.0×10 ⁻¹¹	qSW5
<input checked="" type="checkbox"/>	F8SPD1	F8SPD1_ORYSJ	★	QSW5	Oryza sativa subsp. japonica (Rice)	131	69.0%	174	1.0×10 ⁻¹⁰	qSW5
<input checked="" type="checkbox"/>	F8SPC9	F8SPC9_ORYSI	★	QSW5	Oryza sativa subsp. indica (Rice)	131	69.0%	174	1.0×10 ⁻¹⁰	qSW5
<input checked="" type="checkbox"/>	Q75KY6	Q75KY6_ORYSJ	★	Putative uncharacterized protein OJ1097_A12.5	Oryza sativa subsp. japonica (Rice)	674	100.0%	103	1.7×10 ⁻²	OJ1097_A12.5
<input checked="" type="checkbox"/>	F0Y057	F0Y057_AURAN	★	Expressed protein	Aureococcus anophagefferens (Harmful bloom alga)	590	32.0%	84	2.7	AURANDRAFT_70962
<input checked="" type="checkbox"/>	C1EG84	C1EG84_MICSR	★	Set domain protein	Micromonas sp. (strain RCC299 / NOUM17) (Picoplanktonic green alga)	1,106	42.0%	81	5.9	MICPUN_63548



Domains within the query sequence of 233 residues

1 100 200



Mouse over domain / undefined region for more info; click on it to go to detailed annotation; right-click to save whole protein as PNG image

Transmembrane segments as predicted by the [TMHMM2](#) program (), coiled coil regions determined by the [Coils2](#) program (), segments of low compositional complexity determined by the [SEG](#) program (), Signal peptides determined by the [SignalP](#) program (). Hits only found by BLAST are indicated by [BLAST](#) for hits in the schnipsel database and [PDB](#) for hits against PDB. Regions containing repeats detected by [Prospero](#), but not covered by domains are indicated by [Prospero](#). Disordered regions detected by [DisEMBL](#) (). Intron positions are indicated with vertical lines showing the intron phase and exact position in AA.

You can save the results of your search for easy access in the future by bookmarking this page. It will be available for one month.

The SMART diagram above represents a summary of the results shown below. Domains with scores less significant than established cutoffs are not shown in the diagram. Features are also not shown when two or more occupy the same piece of sequence; the priority for display is given by SMART > PFAM > PROSPERO repeats > Signal peptide > Transmembrane > Coiled coil > Unstructured regions > Low complexity. In either case, features not shown in the above diagram are marked as 'overlap' in the second table below.

Confidently predicted domains, repeats, motifs and features:

Name	Begin	End	E-value
low complexity	24	38	-
low complexity	105	126	-
low complexity	147	160	-
low complexity	176	230	-

These features and domains are not shown in the diagram, either because their scores are less significant than the required threshold, or because they overlap with some other source of annotation:

Name	Begin	End	E-value	Reason
HMG17	28	104	3.13e+03	threshold
Btz	49	143	9.12e+04	threshold
DysFC	69	96	2.06e+03	threshold

DISCOVERED MOTIFS

Motif Overview

[Motif 1](#)

- 6.5e-005
- 4 sites



[Motif 2](#)

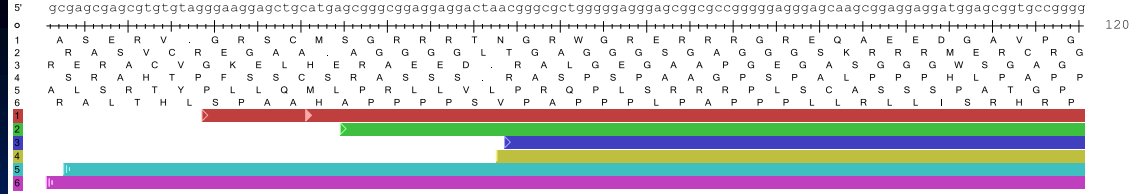
- 7.3e-003
- 2 sites



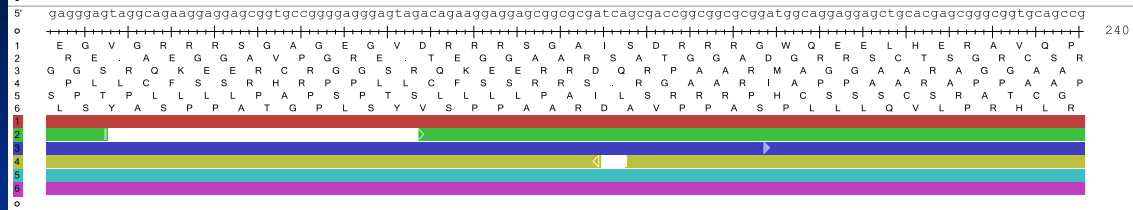
[Motif 3](#)

- 3.3e-003
- 3 sites

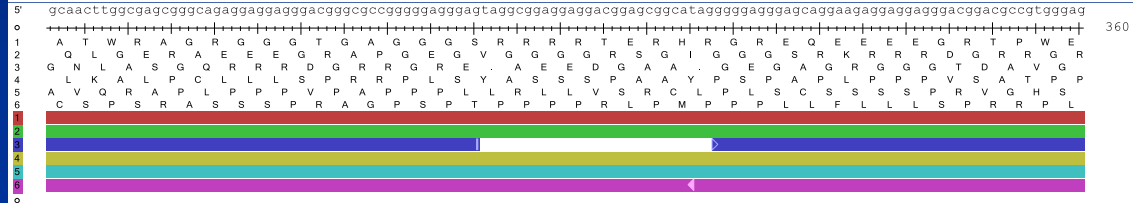




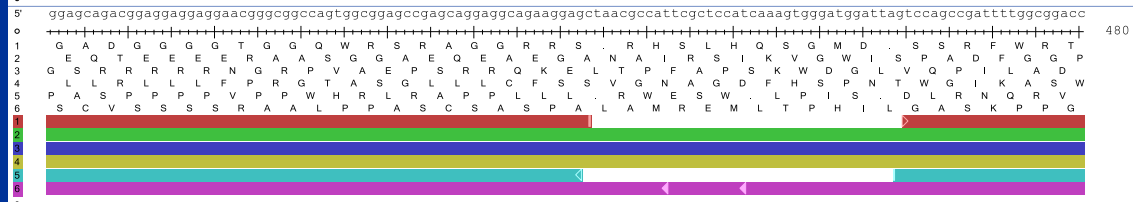
120



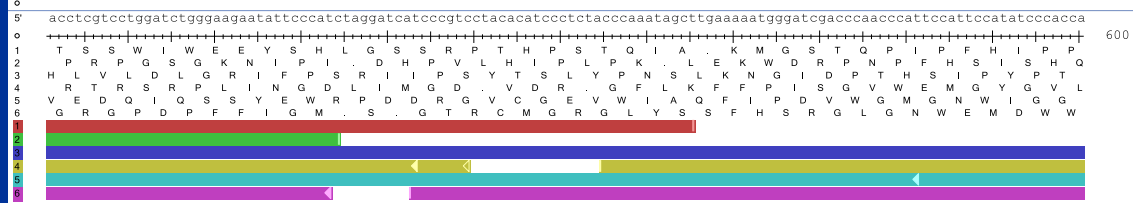
240



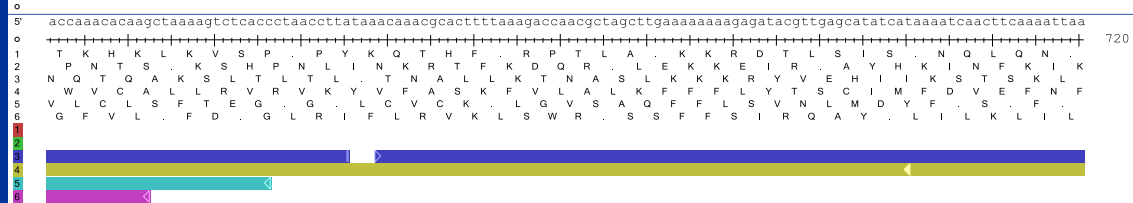
360



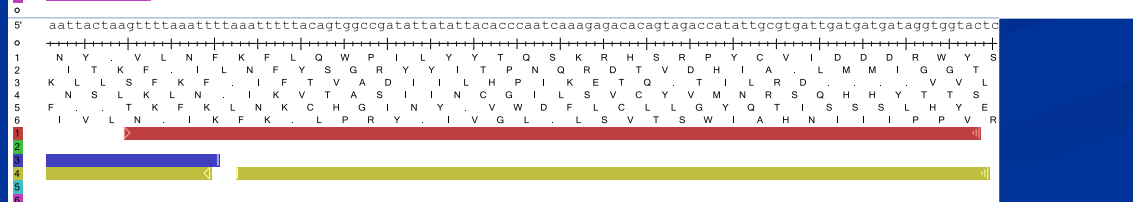
480



600



720



840

已获得CDS区域
六种读码框

小结

- 1、11.2K功能互补序列不含重复序列。
- 2、根据实验数据推测的第一种情况，即含长内含子的预测基因3序列分析显示，与其它物种蛋白同源性极低，无法建立进化树，与表型不能对应，可能不存在这种情况，也可能预测错误，仍需实验验证。首先要通过RACE拿到全长CDNA，然后做表达分析，也可做western验证六种读码框。
- 3、推测该11.2K区域很可能存在两个基因共同起作用，其中一个在籼粳间差异的1.2K, 另一个在上游，需要进行表达验证。



谢谢大家!

敬请指导!