

TABLE S1 Primers used in this study

Purposes	Name of primers	Nucleotide sequences
cDNA sequencing of bvORF18/19/20/21	D-Fw	5'-TGCACCCAGTAGTTGTGCCA-3'
	D-Rv	5'-GCAAGAGAGGATGCCTTAAG-3'
	Gre	5'-GATGCGCGATAATTGTAGCC-3'
Generating hybridization probe of bvORF19 3' UTR	3'-FW	5'-AGCTTGCAAAGCCACTGGGCGA-3'
	3'-RV	5'-GGAACCAAATTAGATTGAATTAACAAGTGG-3'
bvORF16-GFP construction	GFP-ORF16-FW	5'-CCGTGACATGAAATGGAGCTGCGTTG-3'
	GFP-ORF16-RV	5'-GGCCATGGCAGATAGTTCTTTTCCAATTGG-3'
bvORF20-GFP construction	GFP-ORF20-FW	5'-GGGTCGACATGGCATGGTACAGTAAATTC-3'
	GFP-ORF20-RV	5'-GGCCATGGATTTTGCAGACCCAAATAACCC-3'
Amplification of bvORF19 for transgene construction	attB1-ORF19 prom	5'-AAAAGCAGGCTTTAGATCTGCCGTTGCACAACG-3'
	orf19-genomic 3' rv	5'-AGAAAGCTGGGTGTATCTGGGACCTGGATTGAG-3'
Amplification of bvORF20 for transgene construction	attB1-ORF20 prom	5'-AAAAGCAGGCTGTAACAGAGGGTTCAAATTGCGG-3'
	orf20-genomic3'rv	5'-AGAAAGCTGGGTGGTCCTGGATTGAGGGTTAAC-3'
Amplification of bvORF21 for transgene construction	attB1-ORF21 prom	5'-AAAAGCAGGCTGAACCTGAACTGAACTTATTGG-3'
	orf21-genomic3'rv	5'-AGAAAGCTGGGTTACCTGGGTCCCTGGATTAAG-3'
Detection of bialaphos-resistance gene	BAR5	5'-CGAGACAAGCACGGTCAACTTC-3'
	BAR6	5'-AAACCCACGTCATGCCAGTTC-3'

Name of ORFs	Condition of amplification		Nucleotide sequences of primers	NK198						C	Size of PCR products (bp)
	Annealing (°C)	Extension		Anthers		Leaves		Roots			
				+	-	+	-	+	-		
bvORF12	60	1:00	5'-CTGATTTTGGACGGAGCTTGTTTCG-3'							/	630
			5'-TGCATTGTAGAAACACCCGCGTAG-3'								
bvORF13	62	1:30	5'-CCAGGGACAGGGAAGACCAAGAC-3'							/	1100
			5'-AGTCCTCCTTTCCACCCGACAC-3'								
bvORF14	56	0:30	5'-ATCTCCACTTGAAGGGCCAG-3'							/	250
			5'-TTCTCGTCAGACGGACTGAG-3'								
bvORF15	56	0:30	5'-AGTTACCGTGGAGTTACTAGC-3'							/	300
			5'-AGCACAGACTCGTTGCCACT-3'								
bvORF16	52	0:30	5'-AACATCTCCCTAGCCTTCCT-3'							/	870
			5'-CTGAATTCGTTTGCGTATAGT-3'								
bvORF17	56	1:00	5'-CAAGACTTGGTTCAATCAGCC-3'							/	550
			5'-TTCTTTCTCGGCTTCAGCAGC-3'								
bvORF18 /19/20/21	56	0:30	5'-AAGGCATCCTCTCTTGCAAAA-3'							/	360
			5'-TGAATTGCACGTCCTGCTACA-3'								
bvORF22	62	1:30	5'-GTGGCTCTCTCTAAACCGGCTTGT-3'							/	1400
			5'-CATGTTTCAGCCCGACCCACGAA-3'								
bvORF23	53	0:30	5'-CTATTGCGTGATCTTTGTGTTAGAA-3'							/	410
			3'								
bvORF24	56	0:30	5'-TCGAATCTAACGCGGAGACA-3'							/	230
			5'-TGCAGAGGGAGTCAAGTCAG-3'								
bvORF25	54	0:30	5'-ACAGGATTCGCTGGCCTTAA-3'							/	240
			5'-TCAAATGGTCCTCACCAC-3'								

Name of ORFs	Condition of amplification		Nucleotide sequences of primers	NK198							Size of PCR products (bp)
	Annealing (°C)	Extension		Anthers		Leaves		Roots		C	
				+	-	+	-	+	-		
bvORF26	56	1:30	5'-GATGGAAGGTACATGCACAC-3'								100
			5'-CAATGCCACGCCAACTTTCC-3'								
bvORF27	56	0:30	5'-AAGCGTCAGATCCTTAACCC-3'								260
			5'-ACTATTGAGGAAGCTCTGCTGC-3'								
bvORF28	52	0:30	5'-CACCAATTTTAGGGGCTCTA-3'								120
			5'-AAAAATCCAATCCAATAAGTCC-3'								
bvORF29	56	0:30	5'-TTCTCGAACCATATCCCACC-3'								150
			5'-TGTGAAAGTCGAGAGCTAAGG-3'								
bvORF30	56	0:30	5'-ATATTAACCCACGGTCCGG-3'								430
			5'-ATGAGACAGTCGTCCCATAG-3'								
bvORF31	54	1:00	5'-GGATCATACCTGAAGAGTGT-3'								570
			5'-TAAGAAGACCATGCTCTTCC-3'								
bvORF32	56	0:30	5'-TTGAACTTCTAGACCTGGAGT-3'								500
			5'-CACCGAGCTTCTTAAGTAGCATGT-3'								
bvORF33	56	0:30	5'-ACACTTCTTAGGGTGACGAAG-3'								170
			5'-TGTTGAAGCAGTGTGGGGTG-3'								
bvORF34	56	1:00	5'-TGGCAAAGGGGTTTTGACAC-3'								560
			5'-GCAATTCCAGGATCAACATAGCAC-3'								
bvORF35	54	0:30	5'-TCTGATGTATCCACATCATCG-3'								230
			5'-ATTAGATGCATCACGGTCTGG-3'								

FIGURE S1.– RT-PCR analysis of 21 bvORFs. Names of target ORFs, annealing temperatures, extension times, nucleotide sequences of primers, and sizes of PCR products are shown. RNA samples were subjected to reverse transcription with (+) or without (-) reverse transcriptase. Integrity of the PCR reaction was confirmed by control experiments using genomic DNA as templates (C).

CACTTTTGAGCAACTCACAATTTTATATATACATTACAAGTAATTAATAAATAAAGTATTA 100440
TATGGGTAACCTTATACATTGATAAGGGTAACCTAATAAGATAATTTGAGCAACTAATGTTT 100380
TTATATGTACATAACAAGTAAATAATAACACACGTTACATTGACTTGATTACACAAAGGA 100320
 TAACTTATACATTGATAAAGACAACCTAATAAGATAAGTTGAGCAACTAAAGATTTTACAT 100260
TGACTTGGTTAGATAAAGGATAACTTATAGATAGATAAAGGTAACCTAATAAGATAATTTG 100200
AGCAACTAACAATTTTATATATACATTACAAGTAATTAATAAATAAAGTAAAGTATTA 100140
TATGGATAATTAATAAATAAAGTAAATAATTTATATAGAGGAACTATGCAAATTCCTACTA 100080
 ATGTACATTGACTTGGTTTCAGTTAAACTTGTGGAACATGTAATCTTGTAATCCTAGTGG 100020
 AGTAATCTATTTACTTTTAGATTAACCTAAGCATAGTTTGGGATGGCCTATATGATGTTA 99960
 ATAAGACATCACAACCTAATCTCCAACGTTGTTTATATTAATGTGCATGTCATATGAAAAT 99900
 GTCATAGACAAGTTCCAATAAGTCTAGATATGTCAAATTTTCAGTTTACTAAATTATTATT 99840
 GATTTATGTTCTTGTGTTCCAGTAATCTACAATAGAATGGAAGATGAAGAATTTGCTCC 99780
 AACTCCTTGCATTGATGCAACTCCAACCTCCTTGTGTTGATACAACCTCCAACCTAATGCTAC 99720
 TCAAACCTCTAATGCTCAAACAACAAGATCCACATTCCTCCTCGTCCATGCTACACACC 99660
 TAGAGGTTCAAAGAATGGATCCCTTGTGCCCCTCTGAGTTAAACCTACTGTGGGTAT 99600
 R G S K E W I P C C P P E L K P T V G M

 GCCTTTTGATTCTCTTGTGATGGTATTGAGTTTTATAAAGCTTATGCTCGGTTTTGTGG 99540
 P F D S L V D G I E F Y K A Y A R F C G

 TTTTGTGGAAAGATTGGCTACTGAGAAAAAAGATAAGGATGGTCATGTTTACTTGAAGTA 99480
 F V E R L A T E K K D K D G H V Y L K Y

 TATTTATTGTAATAACAAGGATTTAAAGAAGATGGTGAGAGTAAAGCAAAGAGTAAACC 99420
 I Y C N K Q G F K E D G E S K A K S K P

 TATAACATGCTCTAGTTCTCGTAAAGAAGTGTAATCGTGCTGGTTGTCAAGCAAGGAT 99360
 I T C S S S R K R S V N R A G C Q A R I

 AGGTTTGAGAAAACGTAGTGATGGAAAATTCATGGTATATCTTTTTTCATGAATCGCATAA 99300
 G L R K R S D G K F M V Y L F H E S H N

 CCATGTATTTGCCACTCCCAAAGCATGCATTTTCTTAAAAATTCCTCGAACTTGACTCT 99240
 H V F A T P K S M H F L K N S R N L T L
 TGCTCACAAGAAGTTTCATATTTGATAATTCAAGATTGAATGTTGGACCAAACAATCTTT 99180
 A H K K F I F D N S R L N V G P N K S F

 TAGATTGATAAAGAGCATGTAGGAGGATATGAGAATGTAGGGGCGTCATTGGTTGATTT 99120
 R L I K E H V G G Y E N V G A S L V D F

TAAGAACTTCAGTCGAGATGTTAAAGCTTACATACAAGATGTTGATGCCGACATGTTTGT 99060
 K N F S R D V K A Y I Q D V D A D M F V

AAATAATTTCAAAGAAAAGGCAACTAGTAGTGGTGGAGGGTTTTTCTTTGACTATTGTGG 99000
 N N F K E K A T S S G G G F F F D Y C G
 D

ATGAAAATCGACATTTGACTAGAGTTTTTTGGGCGGATGCCATTAGTAGGAAAAACTATT 98940
 *
 E N R H L T R V F W A D A I S R K N Y S

CTCTTTTTGGTGATATGGTATCATTGATACAACCTTTTGATACCAATAAATATTGTATGG 98880
 L F G D M V S F D T T F D T N K Y C M V

TTCTTGCCCCATTTACTGGAGTTGATCATCATGAAAATGTGTTACTTTTTGGTATGGGCC 98820
 L A P F T G V D H H G K C V T F G M G L

TACTTGCAAAGGAAGATATAGAATCTTTCGTTTGGTTGTTTGAATGTTTTTTAAAAGCTA 98760
 L A K E D I E S F V W L F E C F L K A M

TGGGTAATTGTCAACCTACTTGTCTCATTACTGATCAAGATGCAGCAATGAAACAAGCAA 98700
 G N C Q P T C L I T D Q D A A M K Q A I

TTGAAAAAGTTTTCTTTAAGACAATTCATAGACTTTGCGTGTGGCATATCATGAAAAAAG 98640
 E K V F F K T I H R L C V W H I M K K V

TGCCGGTAAAAGTAGGTCCAGATATGTGTAGAACAACGAAGTTTCTTGAGAAATTGAATG 98580
 P V K V G P D M C R T T K F L E K L N A

CTGTTGTTTGGGATAGAGACCTTGAGCCAGATGAATTTGACAAAGGGTGAATTCTGTGA 98520
 V V W D R D L E P D E F D K G W N S V M

TGCGTGAATTTGGCTTAGAAGATGATGGGTGGTTTACTGATATGTTTAAACATAAGACATA 98460
 R E F G L E D D G W F T D M F N I R H M

TGTGGATCCCTTCTTACTTTTCGAAATCTTTTCATGGGTGGTATTTTGAGGTCCACACAGA 98400
 W I P S Y F R N L F M G G I L R S T Q I

TTTCAGAGTCTGAGAACAACCTTTTCACTTTGTTTACAAATGCAAATCTTCTTCTAGTTG 98340
 S E S E N N F F T L F T N A N L L L V E

AGTTATGGTTTTCGGATAGAATCAGCTATGGATGCTCAAAGACATGCCCAAAACAAACTCA 98280
L W F R I E S A M D A Q R H A Q N K L N

ACTCAGATTCTAAGAATTCCATGCCTCGTCTTATTACTCCTCCTTTAGAGAAGCATG 98220
S D S K N S M P R L I T P L P L E K H A

CATCTCTTGTTTACACACACAATATGTTCTACAAATTTAGAGAGAGTTTCAAATGCAA 98160
S L V Y T H N M F Y K F Q R E F Q N A I

TTTTTAATTGTGGGGTTTACAAAGTACAAATAGAGGAAGCTGTTGAGGAGTTTGAAGTTG 98100
F N C G V Y K V Q I E E A V E E F E V A

CAGATAATACAAGGAAGAAAACATATCATGTGACTTTTATTCTGATTCTCATGATTGTT 98040
D N T R K K T Y H V T F I P D S H D C F

TTTGCTCTTGTAAGATGTTTGAATCCATGGGAATATTATGTCGGCATGTGCTTTTTGTGA 97980
C S C K M F E S M G I L C R H V L F V I

TAAAAGGGAAGTTTTTACTGAAATTCAGAGCAACATATATTGCATCGGTGGACTAAAG 97920
K G K F L T E I P E Q H I L H R W T K D

ATGCTTCAAAAAGCCCATTTTTCGACTTTTGTGAGGACTTTGATGGTATAGAAATAAATA 97860
A S K K P I F D F C E D F D G I E I N K

AGAAGAAAAAGTTGTTGGGGATCTTTGGTCGAAATTCCTCTCATGTGTAAGCCTTGTTG 97800
K K K V V G D L W S K F F S C V S L V E

AAAATAACACAGACCATCTTGAGTTATTATTGGAAAGGTTATCTGCTTTTGGAGGAGAAA 97740
N N T D H L E L L L E R L S A F E E E M

TGAAACCTGGAAAAGAAAATGTTGAGCAACAATCTAAAGACAAGCATATTGAGTTGTTTCG 97680
K P G K E N V E Q Q S K D K H I E L F V

TTGGTTCTAATATAGTATCAGGTGGTATACTTCTCCAAACAAGTCTTCAAACAAAGGAA 97620
G S N I V S G G I L P P N K S S N K G S

GTGGTACGGGAAAGAGAAAAGAAAAGTGATCAAGAGATAGCCATTGAAGCAAGCAACAAAA 97560
G T G K R K K S D Q E I A I E A S N K K

AGGAGAGACTGTGTAGATCATGTGGTCAACTTTCAACTCATGATAGTCGTAATTGTCTCTG 97500
 E R L C R S C G Q L S T H D S R N C P D

ATAAAAAGAAAAACCTAGAATGAGGTAAGTTTAGAATAATATATTATTTATACATCACTA 97440
 K K K N L E *

TCAAATGAGAGTTTCTTTTCATTCATTTATTAATAAATTATTTATCACAATTTTTGTTTC 97380
 TTTTGCAGTTACATGGAGGATTGTGATACATGGTGATGACAAGAATCATGATGATCAAAA 97320
 GGGGAGCAAGCAGATTTTACATCACACTTTTTGGGGAATTTTTGTTTGGTTTTATAAAGT 97260
 GTTACCTACAAAAAATAAAGTGTATGTGTTCAAAGGAGCTGGCAAACATTTAAAAATTAT 97200
 GGATTATTCTTTTGAAACGATTTCACTTGAAGGATTTTGATGGCTTAAGCACACTTTTGC 97140
 AGTCTATCTCAAAAATTGTAGGTTTTGTGAAACTATATTCCATGGACACTGCAATTGATA 97080
 TTGTAATTGGAAAGAAGCTTGATTATTCTGTAAGTTTTGCAAGTTATAACTTTTTCTTCTT 97020
 AGGTTGGAAAGAAGTAGATTTTGTAACTTTTTTTTTCTTTTACTATTTGAGGTGACTGGAA 96960
 TCCACCTTTTTGTGTTACCTTTTAGTGCATTCTTTGACTAGGACACTGATTCATTGGTAA 96900
 TTATTATGTAAAAACCCCTATAAACAGTTACTAAAAAGAAGTTATAGTCAACCATACATT 96840
 TACGAGAGAGTATTAATCTTTTTAGTTTGATTTTATGTATTATTTTCAAATATCTTTTTA 96780
 TATGTACATAACAAGTAAATAATTTACACAGTTACATTGACTTGATTACACAAAGGATAA 96720
 CTTATACATTGATAAGGGTAACTAATAAGATAAGTTGAGCAACTAAAGATTTTACATTG 96660
 ACTTGGTTAGCCAAAGGATAACTTATAGATAGATAAGGGTAAATAATATGATAATTTGAG 96600
 CAACTAAACAATTTTATATATACATTACAAGTAATTAATAACAAGAAGTAAAGTATTATA 96540
TGGATAACTTATAGATAGATAAGGGTAACTAATAAGATAATTTGAGCAACTAACAATTTT 96480
ATATATACATTACAAGTAATTAATAACAAGAAGTAAAGTATTATATGGATAACTTATAC 96420
ATTGATAAGGGTAACTAATAAGATAATTTGAGCAACTAACGATTTTATATGTACATAACA 96360
AGTAAATAATAACACACGTTACATTGACTATCTCCTTTTCATCTTTATCAAATAAAAAATA 96300

FIGURE S2.– Nucleotide sequence of *bvMULE-1*. Numbers of nucleotide residues are coordinated with those of BAC clone 5A3. Two open reading frames that exhibit sequence homology to transposases are indicated with putative translation products. The 115-bp and 111-bp repeated sequences are indicated by single and double lines, respectively.

CTCATGACGTCAAGATCTCAAGTCTATAAAAAAAATTGTTTTAAATAAGTCAACCTTTG 11926
 TGCTGTACGTCTTCTTATTTTGGAGTCTACCTTTTGGATTATTTCTGATTGAGTTAAGCTTG 11866
 TATGTATGTTCTCTTCTATTGAGTTTTAATTTATTTATGACTTGTTTAGTAGGTTACTTA 11806
 CTTACTTATGATACGTATTACAATGTCACTCTCGTCTGTTTGAGAAGAAATGACATTGTA 11746
 ACACACATCATAAGTAGTCATTGCATTTGTAATAGCAATCTTAGGTAGAGAGAGAATGCC 11686

 TAGAGAGAGAGAGAGAGAAAACTCTGGAGCGAGCGAAGAGAAGGAGAATGGACAATGGT 11626
 M V
 >ORF-A
 AAGAAGGAGACACCCCCAAGCCAGTAAACCACAACCTAGAGCCTTGAGAACAGCCTTCAT 11566
 R R R H P Q A S K P Q P R A L R T A F I
 RNA recognition motif (pfam accession PF00076)
 AGATTTCTTCTCTCCCAATATTGATACCCAAACAATCCACAACATATTCAGTAGATATGG 11506
D F L P P N I D T Q T I H N I F S R Y G
 TGATCTGGAGGACTTAGTGATACCAGCAAAACTCCGGAAAAACTGTGGGCACAAATACGC 11446
D L E D L V I P A K L R K N C G H K Y A
 ATTCATTAATTTTTTCTCCATGAATGCTTTACTCAATGCGATTAAGCAGGAGAATGGAAG 11386
F I K F F S M N A L L N A I K Q E N G R
 AAGAATGGGAAATTTTTTGGATGCGAGTTAACCTGCAAAATATGACAAACAAGACCCTCC 11326
R M G N F L M R V N P A K Y D K Q D P P
 CCATAAAAACCACTTTCCAAATCCTAAACCAAAATCACAGACAGCCTCAAAAAAACCCGGT 11266
 H K N H F P N P K P N H R Q P Q K N P V
 ACAATATCATCCAGCTTGGAGAGACCACCGATCGTATAAGGATGTCTCGAACCCAAACCA 11206
 Q Y H P A W R D H R S Y K D V S N P N Q
 AATACCAATCCACACTGATGTTCCACCAATCAATCCCTCAACCAAACCTAATACCCGGAA 11146
 I P I H T D V P P I N P S T K P N T R K
 ACCACCTCATCAAACGAATTTATCCTCTTACCTATAGAATCAATCATCCCTAACCAAAT 11086
 P P H Q T N L S S S P I E S I I P N Q I
 CCTTGAACCTCTCAGTACTGACATTGTGAAAGAAATGACAAAGCACCGTAGGATGAGTTC 11026
 L E P L S T D I V K E M T K H R R M S S

TAGGGTCCTTGGGGAAGACACGGAGAGAATAAGGGACCAAGTGGAACTTGTGGAAGTAGA 10966
R V L G E D T E R I R D Q V E L V E L E

GGGCGATCAGATTCTTGCCATCTCAGGGGAGAAAAATGAAGAGATCCTGGAGTTACTGGA 10906
G D Q I L A I S G E K N E E I L E L L E

AAGAAGCGTTATAGCAGTCGCAAACCTTTCATCTCCATCCAAGATTATCCATGAGCATAT 10846
R S V I A V A N S S S P S K I I H E H I

CTTGGCGGAAGGGGTTAACTATCTGAAGATTAAACCCCTTGGGGGAATGCTTCATCTTAT 10786
L A E G V N Y L K I K P L G G M L H L I

CCAGTTCAATTCGGTTGAAGAAAAGGATGACATGATAAAAAGCAAATGGCTTGAACGATG 10726
Q F N S V E E K D D M I K S K W L E R W

GTTCTGGAGCTAAGGGATGTGAATAACGCTAGCACGGCATTATGGAGGGAGATGTGGAT 10666
F L E L R D V N N A S T A L W R E M W I

CACAATTTATGGAGTTCCATTGATCGCATGGAGTTATGAAAATTTTCAGAAAATTGGTTG 10606
T I Y G V P L I A W S Y E N F Q K I G C

TATATTCGGGAGAGTGCTATCGGTGGAATATTCTCGCATGGATTACGCCAGAGTTCAGTT 10546
I F G R V L S V E Y S R M D Y A R V Q L

AATCACAGATTGTCTCTTCAAAGTCAATAACCCCATAGTTTTTTTACGTGGAAGATAAACC 10486
I T D C L F K V N N P I V F Y V E D K P

GTTTAAGATTTTTGTTACAGAAGACTTTGGTCTTGGTCCAAATCATGATCCTCCTGCAAG 10426
F K I F V T E D F G L G P N H D P P A S

TAAAGGTATGCCAAATCCCCTCTTCCATAGATTAGATTCTGATAACTCGAATTCGGAATC 10366
K G M P N P L F H R L D S D N S N S E S

CTCTGATAAAGATCCATTGGATGATGATGATCGTGATAGTGACGACTGGGATCCTCCGGG 10306
S D K D P L D D D D R D S D D W D P P G

AGGGGAAAGGTCACCCCAAAAACCCCTCCCAAACCTCCGAGTTCAATGCATCGGGAAATAC 10246
G E R S P Q K P L P N S E F N A S G N T

TCCAGGAATCTCGGACATTGGAGTTAATAATGAGACAATCATTACTTCGCCTACCAAAGC 10186
P G I S D I G V N N E T I I T S P T K A

TAGTGCCAAGGTCTCTCCAATGCAAAACAAAAACCTCCCATATACCCTAAACCTCCCAA 10126
S A K V S P N A K Q K P P I Y P K P P K

AACTCAACTGAACTTTAATACCCCCACCACGTTCCCCAAGTCTGCTTTGCATTGGGAACTT 10066
T Q L N F N T P P R S P S L L C I G N L

AAATCAACAAAAGTCTCCTCCCAACCACTTGAACTCCAAAAAGCCCCACCTTCACCATC 10006
N Q Q K S S S Q P L E L Q K A P P S P S

GAAAACCTTACCCTTCCCTCCAACAACGAAACTGGGCTCACCTTTTAGCCCTGATCCAAC 9946
K T L P F P P T T K L G S P F S P D P T

CTTTAAATATAATAATCCCCCATCTCCCAAAATAATATAATCAGCCCAATAAGCCCATT 9886
F K Y N N P P I S Q N N I I S P I S P L

GGTCCCCAACCTGCCCAAAATACACAAAACCTCCCTAGTTCTACAAGTCGAAACTCTCC 9826
V P K P A Q N T Q N S P S S T S R N S P

TTTAAAGCCCAGCCTCAATGACCAAAGCTTTCCTTACTACAATCCTCTGATCCACACTGA 9766
L K P S L N D Q S F P Y Y N P L I H T D

TAATTCCTTTGGCCCGCTACTAAGGAAAGCCCAATCAAAATCCCAAACTAAGACACTCTC 9706
N S F G P L L R K A Q S K S Q T K T L S

ATCCTCTCCTTCGACGTCCAGCCCTTCTATCCCCCGGTTTTGAAGACTTCCTTCCTCC 9646
S S P S T S S P S I P P G F E D F L P P

CCCTCTGAAAGCCCATCATGAAAAAGGAGATTACAAAAACGACTGAAGAAAAATAAAGC 9586
P L K A H H E K R R L Q K R L K K N K A

CAAAAACCGCCTCTCCTCCTCCTCCTCCAATCCCCACCTCTTCCTCCCTCTCCCTCCCC 9526
K N R L S S S S S N P P P L P P S P S P

AAACCCGAAAACATCTCATGAGAACACTGCCTCGGAAATTATTGAATTAGGCTTGCAACT 9466
N P K T S H E N T A S E I I E L G L Q L

AGGAATGAAATTCAATGGTGAAC TATCAGATCTACAAGACAAAATTGTTGGAATTTTGTGTC 9406
 G M K F N G E L S D L Q D K I V G I L S

ACGCCAGGAGCAGGACTGGCTTTCCAATGTATAAGTACATCTTATACTCTCAATAAATTG 9346
 R Q E Q D W L S N V *

TTCCATGTTAATCTCGTGAATGTCAGGGGCCTCGGAGCATGGCCTAAAAGAAATGTTCT 9286
 CAAAAAGTTACTACTCCTTCATGACCCCATGATAGTATTCATCCAAGAATCCAAACTGGA 9226
 ATGTATTCCTTCTAAATTGCAAAAATCAATTTGGTGTGATGATGACCTCAGCCTCTGTAT 9166
 CAGTCCATCAAACGGATCCTCTGGAGGATTAATCTCCCTATGGAGACCCCTCAAATTTCA 9106
 TCTGGTTTCCAGTAGAATCGAATCACAATGGATCGCAATGGAAGGAATGGTGGTGAGGGA 9046
 M E G M V V R E
 >ORF-B

AAATTTTCAATGCCTTCTCATAAATATTTATAACTCCTGTGATGCTTCGACTAGATCAGA 8986
 N F Q C L L I N I Y N S C D A S T R S D
 Endonuclease/exonuclease/phosphatase family (pfam accession PF03372)
 CACATGGAACCATATAGAGGATTTTTGCAGAACTCACACTTACCTCTTCTAATAGCGGG 8926
T W N H I E D F C R N S H L P L L I A G

GGATTTCAATGAGGTACTATCTTCCCAAGATCGAGGCAGCCGGATAATAGATGAAACTAG 8866
D F N E V L S S Q D R G S R I I D E T S

TGCCGGAAAATTCAGGCAATTCATAACCAACCTTCATCTTACTGAAATCACACCCTCCAA 8806
A G K F R Q F I T N L H L T E I T P S N

TGGATACTTTACATGGTTTtaggggtcaatcaaaatcaaagctggatagaattcttgcca 8746
G Y F T W F R G Q S K S K L D R I L V Q

ACCAGATTGGATTCTGAAATTCTCATTCCCTTAATGCCTCCATCCTCAAAGGAGTATCTC 8686
P D W I L K F S F L N A S I L K R S I S

GGATCATTGCCCTCTTGTACTGAAGTCGCAATCTAAGGACCGGGGACCGAAACCTTTTCA 8626
D H C P L V L K S Q S K D R G P K P F R

ATTTCTTGACATGTGGCTCACCCACAAGGATTGCCTGATCCTTACTAGGAAAGTATGGGA 8566
 F L D M W L T H K D C L I L T R K V W E

AGATTGGAAGGATTACAATTTAGAGAAGTTTAAAGCTGTCAGAAAAGAGTTGAAAGT 8506
 D S K G F T I S E K F K A V R K E L K V

ATGGAACCAGTCAAAAATTTGGGAATCTAGAAACCAATATCTCTCAATTAGAAGACGAAAT 8446
 W N Q S K F G N L E T N I S Q L E D E I

TCACAAATGGGATACTGTTGCCAACACGAGAAACCTATCGGTTGATGAACTGAGTCTCAG 8386
 H K W D T V A N T R N L S V D E L S L R

ATCCAAGGCCCAACTGGATTTATGGGATTGGATCAAACGCAAGGAAATTCCTGGGCCCA 8326
 S K A Q L D L W D W I K R K E I H W A Q

GAACTCTCGTATAAGTTGGTTGAAGTGTGGGGATAAGAACTCAAAGTTCTTCCATGCCTA 8266
 N S R I S W L K C G D K N S K F F H A Y

TGCATCGATTAGAAGACGGAAGAATAATATCTCTTCCATCACGATCGATGGTGAGACCGT 8206
 A S I R R R K N N I S S I T I D G E T V

CTGTGACCCGGAAAAAATCAAAGCCGAAGCCTCACTCTATTTCCAAAATCTGTTCTCAGA 8146
 C D P E K I K A E A S L Y F Q N L F S E

AGAAACCTTTTCCAGACCAACTTTCTTGAACCTAGCCTTCAAAAAACTCTCATCAATACA 8086
 E T F S R P T F L N L A F K K L S S I Q

ATCCTCGGACCTCACCAAACCTTTCTCACACTCTGAAATAGAAAAAGCAGTAGCATCATG 8026
 S S D L T K P F S H S E I E K A V A S C

TAGCCCTTCAAAATCCCCTGGCCCGGATGGTTTCAATTTTAACTTCATAAAGTCTTCCTG 7966
 S P S K S P G P D G F N F N F I K S S W

GGCAATCATCAAAGAAGACATTTTCTCACTTGTCAATGAATTCTGGCAGTCTGGAACACT 7906
 A I I K E D I F S L V N E F W Q S G T L

ACCAAGGGGTAGTAATGTAGCGTTCATAGCGCTGATCGCCAAGGTGGAAGCCCCCTCAAA 7846
 P R G S N V A F I A L I A K V E A P S N

CTTCAAGGACTTCCGACCCATCAGTATGGTCGGTAGCCTTTACAAGATAATTGCGAAGTT 7786
 F K D F R P I S M V G S L Y K I I A K L

RNA-dependent DNA polymerase (pfam accession PF00078)

GCTTTCCTTCAGGCTGAAAAATGTTATGAACGATCTTATTGGGCCCAACAATCTTCTTT 7726
 L S F R L K N V M N D L I G P Q Q S S F

TATTGAGGGGCGCCAGATCTTGGATAGTGTTTTAATCACTGGCGAGTTATTGGACTCATA 7666
I E G R Q I L D S V L I T G E L L D S Y

CAAAAGTTCCAAGATGGGGGCGAGTAATGTTAAACTGGACTTCCACAAGGCCTTTGACAG 7606
K S S K M G A V M L K L D F H K A F D S

TGTTTCCTGGTCTTTCTTGGATTGGACCATGGATCAAATGGGCTTCCCATTAACATGGCG 7546
V S W S F L D W T M D Q M G F P L T W R

AAAATGGATCTCCTCCTGTGTCTCATCTGCAGCCGCATCTGTCTCCTAAATGGCTCTCC 7486
K W I S S C V S S A A A S V L L N G S P

TTCGACTCCGTTCAAGCTCCAGAGGGGCTCCGTCAAGGAGACCCTCTCTCTCCCTTTCT 7426
S T P F K L Q R G L R Q G D P L S P F L

CTTTGTGTTAGCAGCGGAAGTTTTGAATCTCATGATCAGAAAAGCCACAGAATTGAATAA 7366
F V L A A E V L N L M I R K A T E L N K

ATGGTCAGGTATTGCTATTTGTAAATCGGGTCTATTCTAACTCATCTTCAATTTGCAGA 7306
W S G I A I C K S G P I L T H L Q F A D

TGATACGATAGTATTCTCAACTCCGGATTTGAAGGCGCTCAATAACATCCATAAAACTCT 7246
D T I V F S T P D L K A L N N I H K T L

CATCCTGTTCCAGCTATCCTCAGGCTTGAGATCAACTTCCACAAAAGTGAGATCCTTGG 7186
I L F Q L S S G L Q I N F H K S E I L G

AATCAATACTCCTCAATCTTGGCTTAAAGAAGCGGCAAGGCAATTATTTTGCAGAGTTGG 7126
I N T P Q S W L K E A A R Q L F C R V G

TAATTTCCCGATCACCTACCTGGGCCTTCCAATAGGTGGCAGTTCCGCGAGATTAGCAAC 7066
N F P I T Y L G L P I G G S S A R L A T

ATGGGAACCTCTCTTGGAGAGAATGAGGAAGAAATTGGCCACATGGAAAGAGAAATTACT 7006
W E P L L E R M R K K L A T W K E K L L

CTCGATTGGTGAAGACTCACTTTACTAAAAGCCTCACTCTCGAACCTGCCAATCTATTT 6946
S I G G R L T L L K A S L S N L P I Y F

CATGTCCTTATATCCTATGCCACAAGGAGTTATAGAAAAAATTAATAAAATTCAGAGAAG 6886
 M S L Y P M P Q G V I E K I N K I Q R S

CTTTCTTTGGAGTGGTGGTATGGATAAAAAGGGCTCTATCTATGGTGAAGTGGGAATATGT 6826
 F L W S G G M D K R A L S M V K W E Y V

CCAGCTTCCAAAAGCGTTGGGAGGCTTAAATGTGAGTAACCTTCTGATTAGAAATTTGGG 6766
 Q L P K A L G G L N V S N L L I R N L G

GCTCCTTTGTAAGTGGGTGTGGAGGTATTTTTTCAGAACCAGATTCGCTATGGAGACTATC 6706
 L L C K W V W R Y F S E P D S L W R L S

AATTAAAGCCAAATATAAATACCAGGCGCAAATGAATATGGCTGACATTGCTCCAATAAG 6646
 I K A K Y K Y Q A Q M N M A D I A P I R

ATCAGGTGGTCCCTTGGAGACATCTTTGCAACCATCTCCTAAAACACCAAGCAACAAATGA 6586
 S G G P W R H L C N H L L K H Q A T N E

ACTTCTGAAACAAGGTACCAGGAAAAGAATAGGGAATGGGGAGAATACCTTATTTTGGCA 6526
 L L K Q G T R K R I G N G E N T L F W H

TGACTCTTGGCTGGGCAATTTGCCTCTGAAATTAACCTTCCCAAGACTCTTCCTAATCTC 6466
 D S W L G N L P L K L T F P R L F L I S

AGTTTTACCCATGGCTTCAGTAGCGGAGATGGGTTCTTGGGTGAATTTGGAATGGAAATG 6406
 V L P M A S V A E M G S W V N L E W K W

GAATTTGCCATGGTCCAGAGAATTCAGAAAGAGAGACCGCATTGAATGGGAGCAGCTCCA 6346
 N L P W S R E F R K R D R I E W E Q L Q

ACCTTCCCTCCAGCAAATCTCAGTCCGCCTCAATGAATCAGATGAGTTAATATGGAACCTT 6286
 P S L Q Q I S V R L N E S D E L I W N F

TAGTATGGCTGGTAATTTCTCAGTTCGCTCCTTCTATGAAGAACTTCACAAGCGCTCGAA 6226
 S M A G N F S V R S F Y E E L H K R S K

GCCCTGTCTAGAAAATCTCCCTCAAAGATATGGAAAGGACTTGTTCCCTTCCGAATAGA 6166
 P C L E N L P Q K I W K G L V P F R I E

AATCTTCACTTGGTTATCAGTGCTAGAGAGAATCAATACTAAGAAGAACTAGCTTCTCT 6106
 I F T W L S V L E R I N T K K K L A S L

GAACATTATCCCACCCGCTGAGGTGGGTTGCTCATTATGTAGTTTGGAGCCTGAGGATAT 6046
 N I I P P A E V G C S L C S L E P E D I

TTCGCACCTCTTTTTGTTTTGCCCTTCTCAATGGAGATTTGGGCTTGGTGGTGGGACCT 5986
 S H L F L F C P F S M E I W A W W W D L

TTGGAACCTATCTTGGGTATGGCCAAAATCTCTAAATCTTGCCCTCTCTCAATGGAATTG 5926
 W N L S W V W P K S L N L A L S Q W N C

CCCAAGGAAGGAAAAATTATTCAAAAAATCTGGCTGGCAGCATTTCATTGTGATTATCTG 5866
 P R K E K L F K K I W L A A F I V I I W

GTCAATCTGGAGAGAACGCAATGAGAGAATTTTCAATAAGAAAGAATCATCAGTTTCAGA 5806
 S I W R E R N E R I F N K K E S S V S E

AATCAAAAACCTCATTCTTGTCCGTTTATGTTGGTGGATGAAGCCTTGGAACCTCTCCTT 5746
 I K N L I L V R L C W W M K P W N L S F

CCCGTACACAATTGAAGAAGTCATCAGAATCCCACAATGTCTCTTATGGGGTAGCGCTGT 5686
 P Y T I E E V I R I P Q C L L W G S A V

GCCTCGAAGAAGTAAAACCTCCCATCTCCCCCTCTAATTCAGCTCAGATCTAACCCCC 5626
 P R R S K T S H L P P L I Q L R S N P P

TGACCCTTGTCTCAAGTGGATGGTGGGTTTCACCCCGTTCTCGCCAAAAGAAGGTGCTAG 5566
 D P C L K W M V G F T P F S P K E G A R

AGCAGGAGGCATTTTTGGAGGCTTCCTCAGAGATGAAGTGGGTGTGATCTTATGCTCCTT 5506
 A G G I F G G F L R D E V G V I L C S F

CTCCTGCCCTTTTCCGCCAATGGGTATTAATGAAGTTGCAGTGATTGCAATTCACCGAGC 5446
 S C P F P P M G I N E V A V I A I H R A

TCTGCAAATCTCTCTCAGTGTGCAAATCTAAAAGACCGAGAAATCTCAATTTTCTCTGA 5386
 L Q I S L S V Q N L K D R E I S I F S E

ATCCAGCCAAGCTATCAGTTGGTGCCTCAATCTCTCATCTGGTCCGACTAATCTCTCCTT	5326
S S Q A I S W C L N L S S G P T N L S F	
CCTGTTGAACTTCATCAGATCTACATGCAAAAAGCTCCCTCTTCTGAAGTTTGATTATCT	5266
L L N F I R S T C K K L P L L K F D Y L	
CTCAAGCTGCTCAAGTCAAGTAAAACAGAATGCAATTGGAGAAATCTATGTTTTCTCAGA	5206
S S C S S Q V K Q N A I G E I Y V F S D	
TGTAGTTAGATGGAAAAAGTCTCCCATTTAATTTTGAACCGACCCCCTCCCATGGTATGA	5146
V V R W K K S P I *	
TGTAAGTGGCAGACGAAAGTACCCTTTTAAAAGGATTGTGAAATAAAATGATAAAAAAAAA	5086
AAAAACACACATCATAAGTAACCGATTAAACAAGTCACAAATAGGTTAAAACCTCAATAGA	5026
<u>AGAAAAACAGACATACAAACCTAACTCAACCAAGGATAATCAAAAAGGTAGATTTAAAATGA</u>	4966
<u>GAAAACGCGCAAAAATTGACTTATTTAAAACAATTTTCCAAAAAAAAAATGATATAAAC</u>	4906

FIGURE S3.– Nucleotide sequence of the *BNR-like* element identified in this study. Numbers of nucleotide residues are coordinated with those of BAC clone 33E19. Each pair of palindromic sequences is indicated by single, double, or dotted lines. Putative domains are boxed and shown with pfam descriptions.

TGAGTTTTTCATACTTCAGAACAGTGTGTAATATCCCAAATTTTTATACTATTTTTATAA 79114
AATTAATTTTTGTTATAATAAGATTTTTATATATATATAAATTCTGAAAAATAAAGTAAAAATC 79054
ATAATAAATCAGATTTTTATGAAACTGTTTTATGTATTAAATCAGATTTTTTTTTTAAAA 78994
GAAATTTTCGAAATCAAAATTTTTGAAATTAATCAGATTTTTATCTTTTGAAGAAAAAAA 78934
AATTGGAATTTGATTTTCAGATTTTTCGTCCAAAACGAAAAATAGAGAGAAAAGAAAATT 78874
CTGAAATTATAATTTGAGTTGGTTTTGGAAAAGGATTAGATTTTTGTAAATACTTATCTT 78814
TTAGTGAAACCCTAGATTTACATATATATATATATACCCCCAAAACACCAAAAAAATTCT 78754
CACGTAATACACTTTCCTTCATCTTTTTGGTAAGTTTTAAATTTTCAGATCTAAAATCACCA 78694
TTGTTGTGTTTGAAGGTTTCAACAAAAAAAAAAAAAACTTTTTAAAATCCGATGACTTGGT 78634
CGGAGTCCGGCGTCGGTTTTCTTCTTCTTCTTCTTCTTCTGTTTTCTTCTTCTTCTTCT 78574
TCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 78514
GCGTGCGATGCTTCACAGTACCTCCAAGATTTGATGTTTTTAAAATAAAAAAAGAAAAAA 78454
AAAAAAGAAAAAGGAATTACATGTGCATTAAGTCCACGAGAGATTAGACTTACTGTAA 78394
CATGAAAGTTGAAAAAAAATTTCAAGAAAAAATTTCCCTTTCAATCTGTGACGACGCTGTC 78334
ACGTTTTTTTTCAAGTAAAAGCTTTTTAATTTTTTTTTCTTTTCAACCTTGGTTATTTAA 78274
CTTTGTCAACTATTAAGTATTAACCTTGTAAGCTTTGATTAATTTTTCTTTCGGAATATT 78214
ATATTAAGTGACTAACATTAGGTAATATTATTTGTTGTAGACGGAACTTCGTAGAAGGC 78154
GATTTTTTAGTTGATATTGCTATACTTGGGACGCGTTTGAGGTAATTCACAATGTCCATCA 78094
ACAACTAGAATCCCGTTAGGAATTGTATGTGCTATGTGTAATGCATGTGGTTATGTGATT 78034
CTATATTTATGTTAGATGTATATTATGTGACAGGAATCGTATAGTACTTCTATGACTAA 77974
AATTATTATTATTATTACTATAGTACTATTATCCCTGCGTATAATATATATATGTATCT 77914
ACTGGTATTGTTGTTATGAATTTGGATTTATAATGTACCTAATTATAACCATCGTATTTAA 77854
ATTATGCTAATATTGACAATGTACTTAATGGGTATAACTAGTGTGTTAATGATGTAGCGA 77794
ACGGTATTATAAGTTGACATATATTGTAACCTCTATGAGGCTCTAATGATGGATATATTGG 77734
TTACTACGAATTATGGACGTTGATTAGTATTAAGTGGCTAAGTTGTGAAAAATATTATTT 77674
GAACTAAAGATGTTCCCTGCTAATGTAAATGTGATGTTTGTGATGTGTCACAACTTTTAAAA 77614
TCTATTAATCACGTAAAGTGGAATTTAGGGGAATTATCCTGTGGATTTGGATCCTCCATA 77554
GGTGATGAACAGTACTTGATTTTATTATGATACAACCTTTTATTGTCTTCTCCTAATAC 77494
TATTGGTGCGCATTGCGGATACCCATTAGATTAGTGAGGGGTGTGCACACTAGGGACGCA 77434
CTGCAAAACGATATTGGCCATTGCTCTTAATTTATTGGGTGACGACCCATGTTGAAGTAG 77374
GTGGTTACTGGGATTAGTCCCGCCTACTGACGTTTCGATTCCCTCTAATAATTAATGT 77314
GACGTTGTGTGCATCATATAAACTTGGAGTATATAAATTGATTAAGTGTGTTATGTACTT 77254
GTTTGTATTACGTTATACTTGTCTGCATTGTATTATATTTGTATGTATGTTATAAAAAATT 77194
ATTTTTAAAATACAACATTATTATTATTATTTTTACGGGATGGAGTTGTAATTACTTAGC 77134
TTTCGCTAATTTTTGTGTTTTTGTTTTTCTTTGCTCTTTTCTTATTTATATTGTGCAGGT 77074
TGGTGAAAGGGACTACGTTGCAGGAATGAGTGAATTTATAGTTTTTAAGTTCACCTAGGT 77014
TAATAAACACTACAGTACTGCCATTTGAGATTGTTTAGTTTTGTTTTGGTATAAAGTAGA 76954
CAACTCAGTTTGGTTTATTTTGGATTTTTGGTTTTGTGAGTAAAAATTTTTATTTCTAATAA 76894
TTAGCCTTGCATTTATTTTGAACAATAGATGCGGCGGTAACACCCTAAAGGTTTGGTCAG 76834
ATATTTTAATGAAAAGTGAATATTTTTATAAAGTTTTAGAACCAAAAGTTTTGGGGTGTTAC 76774

AAATTGGTATTCAGAGCTTAGGTTATAATAAATAAATATAAATTAGGAGTTAATAAAATT 76714
TGTTAAATAATTAAGAGTCTAAGATTAATTAAGTAATAAGCATTAAAAATAAATTGTGAG 76654
AGTGGGATCGAAAAGGGTTTATGTGCCTTAATGTGAAAGCACTTATTTTCAGGTGCTTATT 76594
AATTGTTTCTTTTCGTTCTTGTATGTTTAGATGTTGTAAGTACAAGAAAATGTTGTGCTC 76534
GGATGACCGACCATAGGATTTAGAGATTTATATTAAGGTCCGTATACGAAGTTATTTAGT 76474
AGTATTGAAAATTCTCTAAGTTCCTATTTATGTGTTTTAATTATGAATCTTATTACTTGA 76414
CTATTTGAGTTGATTAGAGTTTAAGTTCTTGTATAATTATTGCTTACTTGTCAAATTGA 76354
TGTTATGTGGAAGCATGCAAAGTTGTCTTGAAATGTATGTTATGATGAAATATAGTTTG 76294
GTTGAGTCATGAAATTTTCTTATGATGAAAGTTTGTATTGAGTAAATAAAACGATGTGCA 76234

AGACCATGTATGGGTGAATAGATGTAGTAATACCAATGCTGATGTGCAGGAATTGTATGC 76174
M L M C R N C M Q

AAACCTATCAAGTATACACTAGATCTCGTAGGAGGACTTTTGAACAAATGGCTGAAACCC 76114
T Y Q V Y T R S R R R T F E Q M A E T P

CTGAACAACACTACTTGAGAGATTGAGATCTCTTGAACAACCTTTGCAACGTATGGGTTTAG 76054
E Q L L E R L R S L E Q L S Q R M G L V

TGTTACAAAACCAATTAGGAAATAATGGTGGAGAGGACCCACAAGCTGCTATGGCAAAGA 75994
L Q N Q L G N N G G E D P Q A A M A K K

AGTTAGCAACCCTTAAACCTCCAATCTTTGTAGGAAAGGAAGACCCCTTACTCTTAGAGA 75934
L A T L K P P I F V G K E D P L L L E N

ACTGGCTAAGAGACTTTGATAAGTTATTTCACTGCTACTGGGACACCTGAAGCTCAAAAAG 75874
W L R D F D K L F T A T G T P E A Q K V

TAGACCAAGCTACCTTTTATCTGAGGGAGGATGCAGACACTTGGTGGGAGAGTCAAGGAC 75814
D Q A T F Y L R E D A D T W W E S Q G P

Retrotransposon gag protein (pfam accession PF03732)

CTATTGTTAGAGCTCAGGAAAACCTTTAATTGGAATGCTTTTAAGGTTGCTATTAAGGATA 75754
I V R A Q E N F N W N A F K V A I K D R

GATTTTTCCCTGAACATATTTAGGAGGCAGAAATACAATGAGTTCACTAGATTTAACCAGG 75694
F F P E H I R R Q K Y N E F T R F N Q G

GAGGTACTATGTCTGTGCAAGAGTATGCCAAAAGTTCAATGAGTTAGCTAGATTTTGCC 75634
G T M S V Q E Y A Q K F N E L A R F C P

CTAATGTTGTGCCAGATGAGAGAGCTAAGGCTCAAAAGTTTGAGGATGGTTTAGCATTTA 75574

N V V P D E R A K A Q K F E D G L A F R

GAATTCAGACCAGACTTGGGGGAGCAACTTCTGCAACTTTTCAGGAAGCTTATGCTAAGG 75514

I Q T R L G G A T S A T F Q E A Y A K A

CTTCTAATATTGAGAGGATTTTGAGGCGTGAAGAGGAAGTTATGGGGAGGAATAAGAGAA 75454

S N I E R I L R R E E E V M G R N K R K

AAGACCCACCTAGCAACCAAAATGACCATGGAAATGACAAGAAACCTCGATATGGGGGTA 75394

D P P S N Q N D H G N D K K P R Y G G N

ACAATAATAATGGGGGCAATAATCACACTAATGGTGGTGGTAATTATCAAGGGAATCGTA 75334

N N N G G N N H T N G G G N Y Q G N R S

GCAACTACCAAGGTCAGGGGAGATCAAACCAGCAAGGATCCCGTACCCAGAACCCTACTT 75274

N Y Q G Q G R S N Q Q G S R T Q N P T C

GTAGAAAGTGTAACAAAAGCCACCCAGGATTTACCTGTCAAGGAGACCCAATAACTTGTT 75214

R K C N K S H P G F T C Q G D P I T C Y

Zinc knuckle (pfam accession PF00098)

ATGCTTGTGGAGAGAAAGGGCATAAGGCTAATCAGTGTCCCAAGCGTCAGAATAATGGAC 75154

A C G E K G H K A N Q C P K R Q N N G Q

AAAATGGAAACAATGGGGGAAATAGGAATGGTCATGGGCCTAATCAGAACCAGAATAACA 75094

N G N N G G N R N G H G P N Q N Q N N N

ATAACCGTCCCTACAACAACAACAACCTCTCAAGGTCAAACCTTCGAATGCTCAAGGGGGGA 75034

N R P Y N N N N S Q G Q T S N A Q G G N

ACAATACTCAGCATAATGGTCAGAATAACAATCGAGCAAATGGAGGAAACAACAATCAGA 74974

N T Q H N G Q N N N R A N G G N N N Q N

ACGGCAATGGAAATGGTGCTCGAGGCAACAATGGAAGAATCTATGTTATGAACCAGAATG 74914

G N G N G A R G N N G R I Y V M N Q N E

Retroviral aspartyl protease (pfam accession PF08284)

AAGCAGACACCAACGCCAATGTTGTGACGGGTACTTTCCCTCGTAAACTCTAACCTGCTT 74854

A D T N A N V V T G T F L V N S N P A Y

ACTTGCTTTTTGATTCTGGGGCGTCTCATTCTTTTCATAGCTAGTTCATTTGTTGAAAAGT 74794
 L L F D S G A S H S F I A S S F V E K L

TAGGTCTAAAACCTCAATCTTGTGTCAAACCTTTTCATTACAATACCTTCAGGAGAAGTAG 74734
 G L K P S I L C Q T F I T I P S G E V V

TTCCTTGTAGTTCTCTATACCAAGACATACCCATTACCATATTAGGATCTGATTTGCCGG 74674
 P C S S L Y Q D I P I T I L G S D L P A

CTGATCTTATTCAGTTTGACCTACCCGACTTTGATGTAATATTGGGAATGGATTGGCTTG 74614
 D L I Q F D L P D F D V I L G M D W L A

CTAAGTATAGAGCTAGGATAGAGTGCATACTCAAAGGTGTCTCTTAGGGGGCCAAAGG 74554
 K Y R A R I E C H T Q K V S L R G P K G

GAAATAGAATATCCTATCAAGGAATTGTTTCTAAACCTGGAGTCAGTATTGTGTCAGCCA 74494
 N R I S Y Q G I V S K P G V S I V S A M

TGTCAATCAAACCTATATTAGGAAGGGCTACCCCATATACTTGTGCCATGTGAAGGATG 74434
 S F K T Y I R K G Y P I Y L C H V K D V

TGAGTGTGGAGGATGGAGAGATATCTCAAATACCTGTGGTGAGTGAGTTCCAAGATGTTT 74374
 S V E D G E I S Q I P V V S E F Q D V F

TTCCAGAAGAAATTCCAGGGATGCCGCCAGTGAGAGAAATGGATTTTAAGATTGACCTAG 74314
 P E E I P G M P P V R E M D F K I D L V

TGCCTGGAACCTGGAGCTATTTCTAAGGCACCATATAGGATGGCACCTGCAGAGATGCAAG 74254
 P G T G A I S K A P Y R M A P A E M Q E

AGTTGAAAGTGCAATTGGAGGAATTATTGGAGAAAGGGTACATTAGGCCAAGTGTTCAC 74194
 L K V Q L E E L L E K G Y I R P S V S P

CTTGGGGAGCACCAGTGTTATTTGTTTCGAAAGAAGGATGGAACCTTGAGGTTGTGTATTG 74134
 W G A P V L F V R K K D G T L R L C I D

RNA-dependent DNA polymerase (pfam accession PF00078)

ATTACAGAGAGTTGAATAATGTCACAATAAAGAATAAGTACCCATTGCCTAGGATTGAGG 74074
 Y R E L N N V T I K N K Y P L P R I E D

ATTTATTTGATCAACTTAAGGGTGCTGGAATTTTCTCTAAGATTGATTTGAGGTCTGGGT 74014

L F D Q L K G A G I F S K I D L R S G Y

ATCACCAATTGAGAATTTTCGGAGGAAGATATACCAAAAAACAGCTTTTTCGTACGAGGTATG 73954

H Q L R I S E E D I P K T A F R T R Y G

GGCATTATGAGTTCACAGTGATGCCATTTGGACTTACTAATGCACCTGCAGCATTATGG 73894

H Y E F T V M P F G L T N A P A A F M D

ATCTTATGAATAGAACATTTTCAGCCGTATTTAGATAGATTTGTGGTGGTGTTCATAGATG 73834

L M N R T F Q P Y L D R F V V V F I D D

ATATATTGGTGTATTCGAAGGATAAAGAAGAGCATGAAGGTCATTTAAGGAAAGTTTTGG 73774

I L V Y S K D K E E H E G H L R K V L E

AGATACTTCGAGAGAAAAGGTTGTATGCTAAGTTATCAAAATGTGAGTTTTGGCTTGAGA 73714

I L R E K R L Y A K L S K C E F W L E K

AAGTTGCATTTTTAGGTCATGTGATTTTCGAAGGAAGGTGTTGCTGTAGATCCATCAAAGA 73654

V A F L G H V I S K E G V A V D P S K I

TACAAGCAGTAACAGAATGGGTGAGACCTAGTAATGTGACTGAGATTAGAAGTTTCTTAG 73594

Q A V T E W V R P S N V T E I R S F L G

GACTTGCTGGCTACTATAGGAGGTTTGTGCAAGATTTCTCAAAAAGTAGCTCAACCTTTGA 73534

L A G Y Y R R F V Q D F S K V A Q P L T

CAAATTTGATGAAGAAAACAACCTCGATTTTCAGTGGGATGAGAGGTGTGAGAAAGCTTTTC 73474

N L M K K T T R F Q W D E R C E K A F Q

AGGAATTGAAGCAAAGACTTACTTCAGCACCAGTTTTGACATTACCATCTGGATTAGAAG 73414

E L K Q R L T S A P V L T L P S G L E G

GTTTTGAGGTGTATAGTGACGCTTCTAAGAATGGGTTAGGATGTGTATTGATGCAACATA 73354

F E V Y S D A S K N G L G C V L M Q H S

GTAAGGTGGTAGCATATGCTTCGAGACAACCTTAAGCCTTATGAACAGAATTACCCTACTC 73294

K V V A Y A S R Q L K P Y E Q N Y P T H

ATGATTTAGAGTTAGCTGCTGTAGTATTCGCATTGAAAATTTGGAGGCATTATTTGTATG 73234
 D L E L A A V V F A L K I W R H Y L Y G

GTGTGTCATGTAAGATTTTCACTGATCATAAAAAGTCTGAAATATATATTTACTCAGAAGG 73174
 V S C K I F T D H K S L K Y I F T Q K E

AGTTGAACATGAGACAGAGGAGATGGCTTGAACCTTATTAAGGATTATGATTTAGAGATTT 73114
 L N M R Q R R W L E L I K D Y D L E I L

TGTATCATGAGGGTAAAGCGAATAAAGTTGCTGATGCATTGAGTAGGAAGACTAGTCATT 73054
 Y H E G K A N K V A D A L S R K T S H S

CGATGAACATGATGGTGTATCTGAGAGATTGTGTGAAGATTTTCAGGAGCATGAGTTTAG 72994
 M N M M V L S E R L C E D F R S M S L E

AAGTCATGGAGCAAGGGCAAGTGAAGCTCAATTGAATGCACTATGCGTGCAACCCACCT 72934
 V M E Q G Q V E A Q L N A L C V Q P T L

TATTCGATGAGATTCGAGAGAAGCAAAGTAGTGATGAGTGGATGGTGAAGATAAAGAAAA 72874
 F D E I R E K Q S S D E W M V K I K K M

TGAAAGAAGATGGAGTTGTCATCGAGTTTGACATTGATGAAAATGGTGTGTGAAGTACA 72814
 K E D G V V I E F D I D E N G V V K Y K

AGGGAAGATGGTGTGTTCCCTAAGGATGAGGAGTTAAAAAGAAAGATTTTGGGAAGAAGCTC 72754
 G R W C V P K D E E L K R K I L E E A H

ATAATACTCCATATTCTGTGCATCCTGGAGGAGATAAACTTTATAAGGATTTGAAGCAGC 72694
 N T P Y S V H P G G D K L Y K D L K Q H

ATTTTTGGTGGAAAAACATGAAACGTGAAGTGGCAGAGTTTGTGCAAAGTGTGTTGACGT 72634
 F W W K N M K R E V A E F V A K C L T C

GTCAGAAAGTGAAGATTCAGCATATGAGACCTGGTGAAGTGAATGATGCAACCTTTAGAAGTGC 72574
 Q K V K I Q H M R P G G M M Q P L E V P

Integrase core domain (pfam accession PF00665)

CGAGTTGGAATGGGAGTCTATTTCAATGGATTTTGTGATGGGATTACCACTTACTAAGT 72514
S W K W E S I S M D F V M G L P L T K S

CAGCTAAGAATGCCATATGGGTTATAGTGGATCGATTGACAAAGTCGGCCAGATTTATAG 72454

A K N A I W V I V D R L T K S A R F I A

CAATGAAGGATACATGGAGTATGCAACAGTTGGCTAGTGCATATGTGCGAGAGGTTGTTA 72394

M K D T W S M Q Q L A S A Y V R E V V R

GACTGCATGGAATACCAAAGGATATCGTTTTAGATAGAGACTCGAGATTTTTGTCCAAGT 72334

L H G I P K D I V S D R D S R F L S K F

TTTGGGGGAGGTTACAACAAGCCTTTGGGACATTGCTCAAATTTAGTACAGCTTTCCACC 72274

W G R L Q Q A F G T L L K F S T A F H P

CTGCAACAGATGGACAGACAGAGAGAACAATTCAAACATTGGAGGATATGTTGAGAGCAT 72214

A T D G Q T E R T I Q T L E D M L R A C

GTGTGATAGACTTTGGAGGATCTTGGGATGATTATTTGCCAACTATAGAGTTTTCGTATA 72154

V I D F G G S W D D Y L P T I E F S Y N

ACAACAGTTATCACTCAAGCATAAAGATGGCACCGTATGAAGCATTGTATGGGCGAAAAT 72094

N S Y H S S I K M A P Y E A L Y G R K C

GTAGGAGTCCTTTGTGTTGGAGTGACATAAGTGAGACGATGACTTTAGGGCCTGAGATGA 72034

R S P L C W S D I S E T M T L G P E M I

TTGAAGAAACAACGAAACAAGTTAGGCTTATTCAGGAGCACATGAGGGCAGCTCAAGATA 71974

E E T T K Q V R L I Q E H M R A A Q D R

GACAAAAGGCTTACGCAGATCAGAATAGAAGGGAGATGGAATTTGAGGTTGGGGAGAAGG 71914

Q K A Y A D Q N R R E M E F E V G E K A

CTTTGCTAAAAGTGTACCAACAAGGGGGTCATGAGATTTGGTAGGAAAGGAAAGTTGA 71854

L L K V S P T K G V M R F G R K G K L S

GTCCACGTTACATTGGACCATATGAGATCTTGAACGAATTGGGAAAGTAGCCTATAGAT 71794

P R Y I G P Y E I L E R I G K V A Y R L

TAGCCTTACCAATGGAGTTAGCTAATGTCCATAACGTCTTTCATGTGTCTCAACTTCGAA 71734

A L P M E L A N V H N V F H V S Q L R K

AATATGTCCATGATCCTACCCATATCATTCAACCTGAAACCATTGAACTAGATGAAACCT 71674
Y V H D P T H I I Q P E T I E L D E T L

TATCCTTTGAGCAACGCCAGTTAGGATTCTTGATACCAAAACGAGAAGTACCCGGAACA 71614
S F E Q R P V R I L D T K T R S T R N K
Chromo domain (pfam accession PF00385)

AGGCGGTAAAACCTAGTCAAGGTGTTATGGTCAAGTCAAACCTTCTGAAGAGGCTACTTGGG 71554
A V K L V K V L W S S Q T S E E A T W E

AAGCCGAAGATGATATGAAAAACCGATATCCCGAACCTTCCAGCAGGTACGCTTGAGTT 71494
A E D D M K N R Y P E L S Q Q V R L S F

TCGGGGACGAAACTCTTTAAGGGGGGTAGAATGTGATACTAACTTTTTGTTTGTATATTA 71434
G D E T L *

GTAGCGAGCGATAACGTTAAAGTTCGAGGACGAACTTTCTTTTAAGGGAGAGTAGATGTA 71374
ATATCCCAAATTTTTATAACTATTTTTATAAAATTAATTTTTGTTATAATAAGATTTTTATAT 71314
ATATATAATTCTGAAAAATAAAGTAAATCATAATAAATCAGATTTTTATGAACTGTTT 71254
TTATGTATTAATCAGATTTTTTTTTTAAAGAAATTTTCGAAATCAAAATTTTGAAATTA 71194
ATCAGATTTTTATCTTTTGAAGAAAAAAAAAATTGGAATTTGATTTTCAGATTTTTCGTC 71134
CAAAACGAAAAATAGAGAGAAAAAGAAATTTCTGAAATTTATAATTTGAGTTGGTTTTGGAA 71074
AAGGATTAGATTTTTGTAAATACTTATCTTTTAGTGAAACCCTAGATTTACATATATATA 71014
TATATACCCCCAAAAACACCAAAAAAATTTCTCACGTAATACACTTTCTTCATCTTTTTGGT 70954
AAGTTTTAAATTTTCAGATCTAAATCACCATTGTTGTGTTTGAAGGTTTCAACAAAAAAA 70894
AAAAAACTTTTTAAATCCGATGACTTGGTCGGAGTCCGGCGTCGGTTTTCTTCTTCTCT 70834
TCTTCTTCTGTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 70774
TCCTTCTTCTTCTTCTTATTTTCAATTTGGCTTTGCGTGCGATGCTTTCACAGTACCTCCAAGAT 70714
TTGATGTTTTTAAAAATAAAAAAGAAAAAAAAAAAAAAAAAGAAAAGGAATTACATGTGCATT 70654
AAAGTCCACGAGAGATTAGACTTACTGTAACATGAAAGTTGAAAAAAAAAATTTCAAGAAAA 70594
AATTTCTTTCAATCTGTGACGACGCTGTACGTTTTTTTTTCAAGTAAAAGCTTTTTAAT 70534
TTTTTTTTCTTTTCAACCTTGGTTATTTAACTTTGTCAACTATTAAGTATTAACCTTGTA 70474
AGCTTTGATTAATTTTTCTTTCGGAATATTATATTAAGTGACTAACATTAGGTAATATTA 70414
TTTGTTGTAGACGGAACTTCGTAGAAGGCGATTTTTTAGTTGATATTGCTATACTTGGGA 70354
CGCGTTGAGGTAATTCACAATGTCCATCAACAAGTAAATCCCCTTAGGAATTGTATGT 70294
GCTATGTGTAATGCATGTGGTTATGTGATTCTATATTTATGTTAGATGTATATTATGTGA 70234
CAGGAATCGTATAGTACTTTCTATGACTAAAATTTATTATTATTATTACTATAGTACT 70174
ATTATCCCTGCGTATAATATATATGTATCTACTGGTATTGTTGTTATGAATTTGGATTTA 70114
TAATGTACCTAATTATAACCATCGTATTTAAATTTATGCTAATATTGACAATGTACTTAATG 70054
GGTATAACTAGTGTGTTAATGATGTAGCGAACGGTATTATAAGTTGACATATATTGTAAC 69994
TCTATGAGGCTCTAATGATGGATATATTGGTTACTACGAATTATGGACGTTGATTAGTAT 69934

TAAGTGGCTAAGTTGTGAAAAATATTATTTGAACTAAAGATGTTCCCTGCTAATGTTAATG 69874
TGATGTTTGATGTGTCACAACCTTTTAAAAATCTATTAATCACGTAAAGTGGAAATTGAGG 69814
GAATTATCCTGTGGATTTGGATCCTCCATAGGTGATGAACAGTACTTGATTTTATTATGA 69754
TACAACCTTTTATTGTCTTCCTCCTAATACTATTGGTGCGCATTGCGGATACCCATTAGA 69694

FIGURE S4.– Nucleotide sequence of *bvgypsy-1*. Numbers of nucleotide residues are coordinated with those of BAC clone 33E19. The open reading frame is indicated with its putative translation product. The 1684-bp repeated sequences are underlined. Putative domains are boxed and shown with pfam descriptions.

ORF18	ATGGCGTGGTACAGAAATTC AAGGTTTGTCTACAATGCTTTAAAAC TCAACTTGC GTTCC	60
ORF19	ATGGCGTGGTACAGAAATTC AAGGTTTGTCTACAATGCTTTAAAAC TCAACTTGC GTTCC	60
ORF20	ATGGCGTGGTACAGAAATTC AAGGTTTGTCTACAATGCTTTAAAAC TCAACTTGC GTTCC	60
ORF21	ATGGCGTGGTACAGAAATTC AAGGTTTGTCTACAATGCTTTAAAAC TCAACTTGC GTTCC	60
TK81-O	ATGGCGTGGTACAGAAATTC AAGGTTTGTCTACAATGCTTTAAAAC TCAACTTGC GTTCC	60
ORF18	AAAACATTTGGTACTATTCCA ACTCCAAGAGTTCATT CGAATTCCTCATCTT TGT TTTAC	120
ORF19	AAAACATTTGGTACTATTCCA ACTCCAAGAGTTCATT CGAATTCCTCATCTT TGT TTTAC	120
ORF20	AAAACATTTGGTACTATTCCA ACTCCAAGAGTTCATT CGAATTCCTCATCTT TGT TTTAC	120
ORF21	AAAACATTTGGTACTATTCCA ACTCCAAGAGTTCATT CGAATTCCTCATCTT TGT TTTAC	120
TK81-O	AAAACATTTGGTACTATTCCA ACTCCAAGAGTTCATT CGAATTCCTCATCTT TGT TTTAC	120
ORF18	AATCAATCTACTAATAAGTGT AGTAGGGTTATTTGGGTCTG CAAAATCTGGGTATTTTAA T	180
ORF19	AATCAATCTACTAATAAGTGT AGTAGGGTTATTTGGGTCTG CAAAATCTGGGTATTTTAA T	180
ORF20	AATCAATCTACTAATAAGTGT AGTAGGGTTATTTGGGTCTG CAAAATCTGGGTATTTTAA T	180
ORF21	AATCAATCTACTAATAAGTGT AGTAGGGTTATTTGGGTCTG CAAAATCTGGGTATTTTAA T	180
TK81-O	AATCAATCTACTAA---GTGT AGTAGGGTTATTTGGGTCTG CAAAATCTGGGTATTTTAA T	177
ORF18	GGGTTTAAACATCATCAAGAG ATTAGCTCTTCTCTGGTTTT GCAAGGAGAAATTATCAT	240
ORF19	GGGTTTAAACATCATCAAGAG ATTAGCTCTTCTCTGGTTTT GCAAGGAGAAATTATCAT	240
ORF20	GGGTTTAAACATCATCAAGAG ATTAGCTCTTCTCTGGTTTT GCAAGGAGAAATTATCAT	240
ORF21	GGGTTTAAACATCATCAAGAG ATTAGCTCTTCTCTGGTTTT GCAAGGAGAAATTATCAT	240
TK81-O	GGGTTTAAACATCATCAAGAG ATTAGCTCTTCTCTGGTTTT GCAAGGAGAAACTATCAT	237
ORF18	GGTGATAAAACCGAAGTAAGT GTTGAATCATGGCTGGAAAA ATTCCTTGTTCCAATTGGA	300
ORF19	GGTGATAAAACCGAAGTAAGT GTTGAATCATGGCTGGAAAA ATTACTTCTTC---TTGCA	297
ORF20	GGTGATAAAACCGAAGTAAGT GTTGAATCATGGCTGGAAAA ATTCCTTGTTCCAATTGGA	300
ORF21	GGTGATAAAACCGAAGTAAGT GTTGAATCATGGCTGGAAAA ATTCCTTGTTCCAATTGGA	300
TK81-O	GGTGATAAAACCGAAGTAAGT GTTGAATTTCCGGTGGAAAA ATTACTTCTTGAATTGCA	297
ORF18	C-----TAATCTTGACTTTTGG TATACTTGGTTACCCTCATG <u>TGCACCCAGTAGTT</u>	351
ORF19	GTTGCAC---TAATCTTGA----- TTGCTTACCGTCATG <u>TGCACCCAGTAGTT</u>	342
ORF20	C-----TAATCTTGACTTTTGG TATACTTGGTTACCCTCATG <u>TGCACCCAGTAGTT</u>	351
ORF21	C-----TAATCTTGACTTTTGG TATACTTGGTTACCCTCATG <u>TGCACCCAGTAGTT</u>	351
TK81-O	C-----TAATAATCTCGCATTC TGGTATGATTGCTTTCTTTT ATTT <u>TGCACCCAGTAGTT</u>	351

ORF18	<u>GTGCCA</u> TATACAGGAAGGAAGCATTATGTGCTTATGTCAACAACCTCGTGAGAATGAAATT	411
ORF19	<u>GTGCCA</u> TATACAGGAAGGAAGCATTATGTGCTTATGTCAACAACCTCGTGAGAATGAAAAAT	402
ORF20	<u>GTGCCA</u> TATACAGGAAGGAAGCATTATGTGCTTATGTCAACAACCTCGTGAGAATGAAATT	411
ORF21	<u>GTGCCA</u> TATACAGGAAGGAAGCATTATGTGCTTATGTCAACAACCTCGTGAGAATGAAATT	411
TK81-O	<u>GTGCCA</u> TATACAGGAAGGAAGCATTATGTGATTTTGTCAACAACCTCATGAGAATGAAAAAT	411
	D-Fw	
ORF18	GGAGAAGTTGAGAAGCGGAAAAATACAACCTGCTACACACCCCTGATACTGATAGGGTTAGG	471
ORF19	GGAGAAGTTGAGAAGCGGAAAAATACAACCTGCTACACACCCCTGATACTGAGAGGGTTAGG	462
ORF20	GGAGAAGTTGAGAAGCGGAAAAATACAACCTGCTACACACCCCTGATACTGATAGGGTTAGG	471
ORF21	GGAGAAGTTGAGAAGCGGAAAAATACAACCTGCTACACACCCCTGATACTGATAGGGTTAGG	471
TK81-O	GGAGAATTGAGAAGCGGAAAAATACAACCTGCTACACACCCCTGATACTGAGAGGGTTAGG	471
ORF18	TCAATATTCCAACACATTCTTGAATCACTGGAAAGAGAGATTAATCACCATGAACTCGAA	531
ORF19	TCTATATTCCAACACATTATTGAATCACTGGAAAGAGAGATTAATCACCATGAACTCGAA	522
ORF20	TCAATATTCCAACACATTCTTGAATCACTGGAAAGAGAGATTAATCACCATGAACTCGAA	531
ORF21	TCAATATTCCAACACATTCTTGAATCACTGGAAAGAGAGATTAATCACCATGAACTCGAA	531
TK81-O	TCTATATTCCAACACATTCTTGAATCACTGGAAAGAGAGATTAATCACCATGAACTCGAA	531
ORF18	CTCGAACTCGAA-----AGAGATGAAACTTTCAAGGAGAAAACCATTGGAAGGAGGAG	585
ORF19	CTCGAA-----AGAGATGAAACTTTCAAGGAGAAAACCATTGGAAGGAGGAG	570
ORF20	CTCGAA-----AGAGATGAAACTTTCAAGGAGAAAACCATTGGAAGGAGGAG	579
ORF21	CTCGAACTCGAA-----AGAGATGAAACTTTCAAGGAGAAAACCATTGGAAGGAGGAG	585
TK81-O	CTCGAACTCGAACTCGAAAGAGATGAAACTTTCAAGGAGAAAACCATTGGAAGGAGGAG	591
ORF18	ACAGTTGATGATAAAGATAGTAGGAAGAAGCATAAGTGGGGCTAAGATAACTACTAACCAT	645
ORF19	ACAGTTGATGATAAAGATAGTAGGAAGAAGCATAAGTGGGGCTAAGATAACTACTAACCAT	630
ORF20	ACAGTTGATGATAAAGATAGTAGGAAGAAGCATAAGTGGGGCTAAGATAACTACTAACCAT	639
ORF21	ACAGTTGATGATAAAGATAGTAGGAAGAAGCATAAGTGGGGCTAAGATAACTACTAACCAT	645
TK81-O	ACAGATCATGATAAAGATAGTAGGAAGAAGCATAAGTGGGGCTAAGATAACTACTAACCAT	651
ORF18	TTGGAAGGGATGAATTGGGAAATTTTCGTTGTTGATAAACCGTTGGTTGAGTCCAGTTAT	705
ORF19	TTGGAAGGGTGAATTGGGAAATTTTCGTTGTTGATAAACCGTTGGTTGAGTCCAGTTGT	690
ORF20	TTGGAAGGGATGAATTGGGAAATTTTCGTTGTTGATAAACCGTTGGTTGAGTCCAGTTAT	699
ORF21	TTGGAAGGGATGAATTGGGAAATTTTCGTTGTTGATAAACCGTTGGTTGAGTCCAGTTAT	705
TK81-O	---GAAGGGATGAATTGGGAAATTTTCGTTGTCGATAAACCGTTGGTTGAGTCCAGTTGT	708

ORF18 TTATTAGGTGGGAAGATTGTTGTTTACACCGGATTGCTCAACCATT-GCAACTCTGATG 763
 ORF19 TTATTTGATGGGAAGATTGTTGTTTACACCGGATTGCTCAACCATTT-CAACTCTGATG 748
 ORF20 TTATTAGGTGGGAAGATTGTTGTTTACACCGGATTGCTCAACCATT-GCAACTCTGATG 757
 ORF21 TTATTAGGTGGGAAGATTGTTGTTTACACCGGATTGCTCAACCATT-GCAACTCTGATG 763
 TK81-0 ATATTTGGTGGGAAGATTGTTGTTTACACTGGATTGCTCAACCATTTG-ATCTCTGATG 766

Intron 1

ORF18 CTGAATTGGCTACAATTATCGCGCATCAGGTTGGGCATGCTGTGGCTCGACATGAGGCAG 823
 ORF19 CTGAATTGGCTACAATTATCGCGCATCAGGTTGGGCATGCTGTGGCTCGACATGAGGCAG 808
 ORF20 CTGAATTGGCTACAATTATCGCGCATCAGGTTGGGCATGCTGTGGCTCGACATGAGGCAG 817
 ORF21 CTGAATTGGCTACAATTATCGCGCATCAGGTTGGGCATGCTGTGGCTCGACATGAGGCAG 823
 TK81-0 CTGAATTGGCTACAATTATCGCGCATCAGGTTGGGCATGCTGTGGCTCGACATGAGGCAG 826

Gre

ORF18 AGGATTGCAGACGATTTTCTGGTTGTTAATA---TCCCTCAACGTGATATTATTTAAAA 880
 ORF19 AGCATTGGACAGCATTGTTCTGGTGGTCAATGTTAGGGTTCACGTGACATTATTTGAAA 868
 ORF20 AGGATTGCAGACGATTTTCTGGTTGTTAATA---TCCCTCAACGTGATATTATTTAAAA 874
 ORF21 AGGATTGCAGACGATTTTCTGGTTGTTAATA---TCCCTCAACGTGATATTATTTAAAA 880
 TK81-0 AGCATTGGACAACATTGTTGTGGTCGATACTGTTAGTGATATACATGACAATATTTCAAT 886

ORF18 TTCTATTTACTGAGCCTGAATCTGCCAATGCAAGATCAAACTACTCTTAAGGCATCCTC 940
 ORF19 TTCTATTTACTGCGCCTGAATTTGCCAATGCAAGATCAAACTACTCTTAAGGCATCCTC 928
 ORF20 TTCTATTTACTGAGCCTGAATCTGCCAATGCAAGATCAAACTACTCTTAAGGCATCCTC 934
 ORF21 TTCTATTTACTGAGCCTGAATCTGCCAATGCAAGATCAAACTACTCTTAAGGCATCCTC 940
 TK81-0 ATCTATTTACTGCGCCTGAATTTGCCAATGCAATATCAAACTACTCTCAAGGCATCCTC 946

Intron 2

ORF18 TCTTGCAAAAAGTTTGGGAAGATTATTCAGGCTAGAGCTCCACAATTACTGCCACGAACTA 1000
 ORF19 TCTTGCAAAAAGTTTGGGAAGATTATTCAGGCTAGATTTCATCAATTACTGCCACGAACTA 988
 ORF20 TCTTGCAAAAAGTTTGGGAAGATTATTCAGGCTAGAGCTCCACAATTACTGCCACGAACTA 994
 ORF21 TCTTGCAAAAAGTTTGGGAAGATTATTCAGGCTAGAGCTCCACAATTACTGCCACGAACTA 1000
 TK81-0 TCTTGCAAAAAGTTTGGGAAGATTATTCAGGCTAGATTTCATCAATTACTGCCACGAACTA 1006

D-Rv

ORF18 TCT---GCTTGTCCTTGTGGATTGTTTCCCTCGGTGTTTATTCTTTATTATGGTCGGA 1057
 ORF19 CCTTGCAGATTGGGCTTTGTTGGATTGTCTTCCCTGGTGTATTCTTTATTTTGGTCGGA 1048
 ORF20 TCT---GCTTGTCCTTGTGGATTGTTTCCCTCGGTGTTTATTCTTTATTATGGTCGGA 1051
 ORF21 TCT---GCTTGTCCTTGTGGATTGTTTCCCTCGGTGTTTATTCTTTATTATGGTCGGA 1057
 TK81-0 CCTTGCACCTTGGGCTTTCTTGGATTGTCTTCCCTGGTGTATTCTTTATTTTGGTCGGA 1066

```

ORF18      AGGAAATAGAAGCAGATCACATTGGAGTGCTTCTGATGGCTTCTGCTGGATACGACCCGC 1117
ORF19      AGGAAATAGAAGCAGATCACATTGGAGTGCTTCTGATGGCTTCTGCTGGATACGACCCGC 1108
ORF20      AGGAAATAGAAGCAGATCACATTGGAGTGCTTCTGATGGCTTCTGCTGGATACGACCCGC 1111
ORF21      AGGAAATAGAAGCAGATCACATTGGAGTGCTTCTGATGGCTTCTGCTGGATACGACCCGC 1117
TK81-O     AGGAAATAGAAGCAGATCACATTGGAGTGCTTCTGATGGCTTCTGCTGGATACGACCCGC 1126

ORF18      GAGTTGCACCTCAAGTATATGACAAGCTTGCAAAGCCACTGGGCGACTGGAAGTGTTTAG 1177
ORF19      GAGTTGCACCTCAAGTATATGACAAGCTTGCAAAGCCACTGGGCGACTGGAAGTGTTTAG 1168
ORF20      GAGTTGCACCTCAAGTATATGACAAGCTTGCAAAGCCACTGGGCGACTGGAAGTGTTTAG 1171
ORF21      GAGTTGCACCTCAAGTATATGACAAGCTTGCAAAGCCACTGGGCGACTGGAAGTGTTTAG 1177
TK81-O     GAGTTGCACCTCAAGTATATGACAAGCTTGCAAAGCCACTGGGCGACTGGAAGTGTTTAG 1186

ORF18      CAACTCATCCATTTGCAAGAATGAGAGCAAAGTTGTTAGCTCGAGCTGATGTTATGAAGG 1237
ORF19      CAACTCATCCATTTGCAAGAATGAGAGCAAAGTTGTTAGCTCGAGCTGATGTTATGAAGG 1228
ORF20      CAACTCATCCATTTGCAAGAATGAGAGCAAAGTTGTTAGCTCGAGCTGATGTTATGAAGG 1231
ORF21      CAACTCATCCATTTGCAAGAATGAGAGCAAAGTTGTTAGCTCGAGCTGATGTTATGAAGG 1237
TK81-O     CAACTCATCCATTTGCAAGAATGAGAGCAAAGTTGTTAGCTCGAGCTGATGTTATGAAGG 1246

ORF18      AAGCAGATAAGATATACAATGAAGTTGTAGCAGGACGTGCAATTCAAGGTCTTCAGTAA 1296
ORF19      AAGCAGATAAGATATACAATGAAGTTGTAGCAGGACGTGCAATTCAAGGTCTTCAGTAA 1287
ORF20      AAGCAGATAAGATATACAATGAAGTTGTAGCAGGACGTGCAATTCAAGGTCTTCAGTAA 1290
ORF21      AAGCAGATAAGATATACAATGAAGTTGTAGCAGGACGTGCAATTCAAGGTCTTCAGTAA 1296
TK81-O     AAGCAGATAAGATATACAATGAAGTTGTAGCAGGACGTGCAATTCAAGGTCTTCAGTAA 1305

```

FIGURE S5.– Sequence alignment of *bvORF18* (ORF18), *bvORF19* (ORF19), *bvORF20* (ORF20), *bvORF21* (ORF21), and *bvORF20L* (TK81-O). Hyphens indicate gaps inserted for maximum matching. Residues of nucleotide sequences are numbered from the translational initiation codon. Positions of introns are shown with black triangles, but the intronic sequences are not shown. Exon/intron boundaries have been experimentally confirmed (H. Matsuhira, T. Mikami and T. Kubo, manuscript in preparation). Primer sequences are underlined. Nucleotide residues corresponding to Site 1 and Site 2 in Figure 4 are shown by red and blue letters, respectively. 5'-CTCGAA-3' repeated sequences are indicated by purple letters.

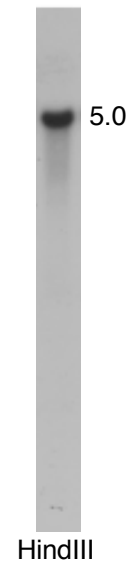


FIGURE S6.– Gel blot analyses using bvORF16 sequence as a probe. Total cellular DNA of NK–198 was used. Size of signal band is given in kbp. DNA fragment for the hybridization probe was generated by PCR using a pair of primers (5'-TGTGTATGCTGTTCTGGTTGA -3' and 5'-AACATCTCCCTAGCCTTCCT -3') and a BAC clone DNA as a template.

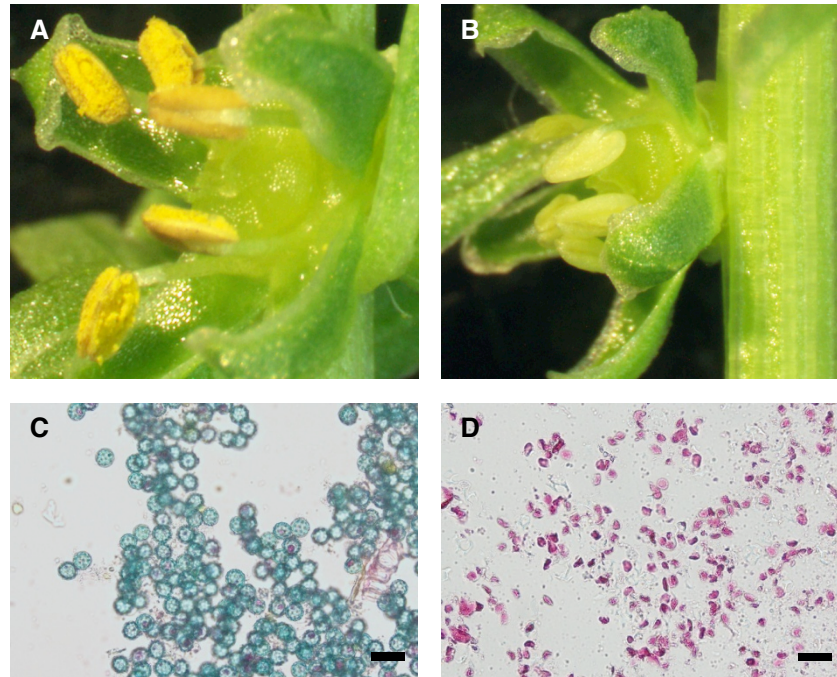


FIGURE S7.– Comparison of anther morphology and anther content between two sugar-beet plants derived from the 14 F1 plants. Panels A and C show photographs taken from a plant having the biaphos-resistance gene. Panels B and D show photographs taken from a plant missing the biaphos-resistance gene. A and B, anther morphology. C and D, images of Alexander's staining (scale bars; 20 μ m).

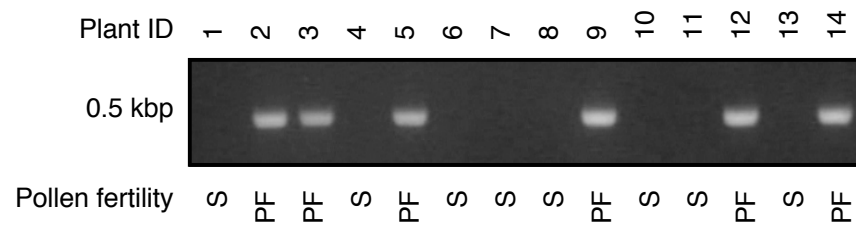


FIGURE S8.– Co-segregation analysis between the bialaphos-resistance gene and partial fertility. Agarose gel electrophoresis of PCR products resulting from amplification using primers targeting the bialaphos-resistance gene. Size of the amplicon is shown on the left. Plant ID and pollen fertility is shown above and below the photograph, respectively. PF and S indicate partial fertility and complete sterility, respectively.

File S1. List of gene sequences used for phylogenetic analysis.

Name of genes	Label in tree
AtPPR_1g01970*	AT1g01970
AtPPR_1g02060	AT1g02060
AtPPR_1g02150	AT1g02150
AtPPR_1g02370	AT1g02370
AtPPR_1g02420	AT1g02420
AtPPR_1g03100	AT1g03100
AtPPR_1g03560	AT1g03560
AtPPR_1g05600	AT1g05600
AtPPR_1g05670	AT1g05670
AtPPR_1g06270	AT1g06270
AtPPR_1g06580	AtRFL1
AtPPR_1g06710	AT1g06710
AtPPR_1g07590	AT1g07590
AtPPR_1g07740	AT1g07740
AtPPR_1g08610	AT1g08610
AtPPR_1g09680	AT1g09680
AtPPR_1g09820	AT1g09820
AtPPR_1g09900	AT1g09900
AtPPR_1g10270	AT1g10270
AtPPR_1g10910	AT1g10910
AtPPR_1g11630	AT1g11630
AtPPR_1g11710	AT1g11710
AtPPR_1g11900	AT1g11900
AtPPR_1g12300	AtRFL2
AtPPR_1g12620	AtRFL3
AtPPR_1g12700	AtRFL4
AtPPR_1g12770	AT1g12770
AtPPR_1g13040	AT1g13040
AtPPR_1g13630	AT1g13630
AtPPR_1g13800	AT1g13800
AtPPR_1g15480	AT1g15480
AtPPR_1g16830	AT1g16830
AtPPR_1g18900	AT1g18900
AtPPR_1g19290	AT1g19290
AtPPR_1g19520	AT1g19520
AtPPR_1g20300	AT1g20300
AtPPR_1g22960	AT1g22960
AtPPR_1g26460	AT1g26460
AtPPR_1g26500	AT1g26500
AtPPR_1g28020	AT1g28020
AtPPR_1g30610	AT1g30610
AtPPR_1g31790	AT1g31790
AtPPR_1g31840	AT1g31840
AtPPR_1g43010	AT1g43010
AtPPR_1g51965	AT1g51965
AtPPR_1g52620	AT1g52620
AtPPR_1g52640	AT1g52640
AtPPR_1g53330	AT1g53330
AtPPR_1g55630	AT1g55630
AtPPR_1g55890	AT1g55890
AtPPR_1g60770	AT1g60770
AtPPR_1g61870	AT1g61870

AtPPR_1g62350	AT1g62350
AtPPR_1g62590	AtRFL5
AtPPR_1g62670	AtRFL6
AtPPR_1g62680	AtRFL7
AtPPR_1g62720	AtRFL8
AtPPR_1g62910	AtRFL9
AtPPR_1g62930	AtRFL11
AtPPR_1g63070	AtRFL12
AtPPR_1g63080	AtRFL13
AtPPR_1g63130	AtRFL14
AtPPR_1g63150	AtRFL15
AtPPR_1g63330	AtRFL16
AtPPR_1g63400	AtRFL17
AtPPR_1g64100	AtRFL18
AtPPR_1g64580	AtRFL19
AtPPR_1g64585	AT1g64585
AtPPR_1g66345	AT1g66345
AtPPR_1g68980	AT1g68980
AtPPR_1g69290	AT1g69290
AtPPR_1g71060	AT1g71060
AtPPR_1g71210	AT1g71210
AtPPR_1g73400	AT1g73400
AtPPR_1g73710	AT1g73710
AtPPR_1g74580	AT1g74580
AtPPR_1g74750	AT1g74750
AtPPR_1g74850	AT1g74850
AtPPR_1g74900	AT1g74900
AtPPR_1g76280	AT1g76280
AtPPR_1g77360	AT1g77360
AtPPR_1g77405	AT1g77405
AtPPR_1g79080	AT1g79080
AtPPR_1g79490	AT1g79490
AtPPR_1g79540	AT1g79540
AtPPR_1g80150	AT1g80150
AtPPR_1g80270	AT1g80270
AtPPR_1g80550	AT1g80550
AtPPR_1g80880	AT1g80880
AtPPR_2g01360	AT2g01360
AtPPR_2g01390	AT2g01390
AtPPR_2g01740	AT2g01740
AtPPR_2g01860	AT2g01860
AtPPR_2g02150	AT2g02150
AtPPR_2g06000	AT2g06000
AtPPR_2g13420	AT2g13420
AtPPR_2g15630	AT2g15630
AtPPR_2g15820	AT2g15820
AtPPR_2g15980	AT2g15980
AtPPR_2g16880	AT2g16880
AtPPR_2g17033	AT2g17033
AtPPR_2g17140	AT2g17140
AtPPR_2g17525	AT2g17525
AtPPR_2g17670	AT2g17670
AtPPR_2g18520	AT2g18520
AtPPR_2g18940	AT2g18940

AtPPR_2g19280	AT2g19280
AtPPR_2g20710	AT2g20710
AtPPR_2g26790	AT2g26790
AtPPR_2g27800	AT2g27800
AtPPR_2g28050	AT2g28050
AtPPR_2g30100	AT2g30100
AtPPR_2g30780	AT2g30780
AtPPR_2g31400	AT2g31400
AtPPR_2g32230	AT2g32230
AtPPR_2g32630	AT2g32630
AtPPR_2g35130	AT2g35130
AtPPR_2g36240	AT2g36240
AtPPR_2g37230	AT2g37230
AtPPR_2g38420	AT2g38420
AtPPR_2g39230	AT2g39230
AtPPR_2g40240	AT2g40240
AtPPR_2g41720	AT2g41720
AtPPR_2g48000	AT2g48000
AtPPR_3g02490	AT3g02490
AtPPR_3g02650	AT3g02650
AtPPR_3g04130	AT3g04130
AtPPR_3g04760	AT3g04760
AtPPR_3g06430	AT3g06430
AtPPR_3g06920	AT3g06920
AtPPR_3g07290	AT3g07290
AtPPR_3g09060	AT3g09060
AtPPR_3g09650	AT3g09650
AtPPR_3g13150	AT3g13150
AtPPR_3g13160	AT3g13160
AtPPR_3g14580	AT3g14580
AtPPR_3g15200	AT3g15200
AtPPR_3g15590	AT3g15590
AtPPR_3g16010	AT3g16010
AtPPR_3g16710	AtRFL20
AtPPR_3g16890	AT3g16890
AtPPR_3g18020	AT3g18020
AtPPR_3g18110	AT3g18110
AtPPR_3g22470	AtRFL21
AtPPR_3g22670	AT3g22670
AtPPR_3g23020	AT3g23020
AtPPR_3g25210	AT3g25210
AtPPR_3g29290	AT3g29290
AtPPR_3g42630	AT3g42630
AtPPR_3g46610	AT3g46610
AtPPR_3g46870	AT3g46870
AtPPR_3g48250	AT3g48250
AtPPR_3g48810	AT3g48810
AtPPR_3g49240	AT3g49240
AtPPR_3g49730	AT3g49730
AtPPR_3g53170	AT3g53170
AtPPR_3g53700	AT3g53700
AtPPR_3g54980	AT3g54980
AtPPR_3g56030	AT3g56030
AtPPR_3g59040	AT3g59040

AtPPR_3g60050	AT3g60050
AtPPR_3g60960	AT3g60960
AtPPR_3g60980	AT3g60980
AtPPR_3g61360	AT3g61360
AtPPR_3g61520	AT3g61520
AtPPR_3g62470	AT3g62470
AtPPR_3g62540	AT3g62540
AtPPR_4g01400	AT4g01400
AtPPR_4g01570	AT4g01570
AtPPR_4g01990	AT4g01990
AtPPR_4g02820	AT4g02820
AtPPR_4g04790	AT4g04790
AtPPR_4g11690	AT4g11690
AtPPR_4g14190	AT4g14190
AtPPR_4g16390	AT4g16390
AtPPR_4g17618	AT4g17618
AtPPR_4g17910	AT4g17910
AtPPR_4g18975	AT4g18975
AtPPR_4g19440	AT4g19440
AtPPR_4g19900	AT4g19900
AtPPR_4g20090	AT4g20090
AtPPR_4g20740	AT4g20740
AtPPR_4g21170	AT4g21170
AtPPR_4g21190	AT4g21190
AtPPR_4g21705	AT4g21705
AtPPR_4g21880	AT4g21880
AtPPR_4g26680	AT4g26680
AtPPR_4g28010	AT4g28010
AtPPR_4g30825	AT4g30825
AtPPR_4g31850	AT4g31850
AtPPR_4g34830	AT4g34830
AtPPR_4g35850	AT4g35850
AtPPR_4g36680	AT4g36680
AtPPR_4g38150	AT4g38150
AtPPR_4g39620	AT4g39620
AtPPR_5g01110	AT5g01110
AtPPR_5g02830	AT5g02830
AtPPR_5g02860	AT5g02860
AtPPR_5g04810	AT5g04810
AtPPR_5g06400	AT5g06400
AtPPR_5g08315	AT5g08315
AtPPR_5g09450	AT5g09450
AtPPR_5g10690	AT5g10690
AtPPR_5g11310	AT5g11310
AtPPR_5g12100	AT5g12100
AtPPR_5g13770	AT5g13770
AtPPR_5g14080	AT5g14080
AtPPR_5g14770	AT5g14770
AtPPR_5g14820	AT5g14820
AtPPR_5g15010	AT5g15010
AtPPR_5g15280	AT5g15280
AtPPR_5g15980	AT5g15980
AtPPR_5g16420	AT5g16420
AtPPR_5g16640	AtRFL23

AtPPR_5g18390	AT5g18390
AtPPR_5g18475	AT5g18475
AtPPR_5g18950	AT5g18950
AtPPR_5g21222	AT5g21222
AtPPR_5g24830	AT5g24830
AtPPR_5g25630	AT5g25630
AtPPR_5g27270	AT5g27270
AtPPR_5g27460	AT5g27460
AtPPR_5g28460	AT5g28460
AtPPR_5g38730	AT5g38730
AtPPR_5g39710	AT5g39710
AtPPR_5g39980	AT5g39980
AtPPR_5g40400	AT5g40400
AtPPR_5g41170	AtRFL24
AtPPR_5g42310	At_CRP1
AtPPR_5g43820	AT5g43820
AtPPR_5g46100	AT5g46100
AtPPR_5g46580	AT5g46580
AtPPR_5g46680	AT5g46680
AtPPR_5g47360	AT5g47360
AtPPR_5g48730	AT5g48730
AtPPR_5g50280	AT5g50280
AtPPR_5g55840	AT5g55840
AtPPR_5g57260	AT5g57260
AtPPR_5g59900	AT5g59900
AtPPR_5g60960	AT5g60960
AtPPR_5g61370	AT5g61370
AtPPR_5g61400	AT5g61400
AtPPR_5g61990	AT5g61990
AtPPR_5g62370	AT5g62370
AtPPR_5g64320	AT5g64320
AtPPR_5g65560	AT5g65560
AtPPR_5g65820	AT5g65820
AtPPR_5g66635	AT5g66635
AtPPR_5g67570	AT5g67570
AtPPR_1g62910	AtRFL10
AtPPR_4g26800	AtRFL22
AtPPR_1g12770	AtRFL25
AtPPR_1g64585	AtRFL26
Glyma01g07140 ^{*2}	GmRFL1
Glyma01g07300	GmRFL2
Glyma02g09530	GmRFL3
Glyma05g28430	GmRFL4
Glyma0679s00210	GmRFL5
Glyma07g11290	GmRFL6
Glyma07g11410	GmRFL7
Glyma07g27410	GmRFL8
Glyma08g05770	GmRFL9
Glyma09g07250	GmRFL10
Glyma09g07290	GmRFL11
Glyma09g07300	GmRFL12
Glyma09g28360	GmRFL13
Glyma09g30160	GmRFL14
Glyma09g30500	GmRFL15

Glyma09g30530	GmRFL16
Glyma09g30620	GmRFL18
Glyma09g30640	GmRFL19
Glyma09g30680	GmRFL20
Glyma09g30720	GmRFL21
Glyma09g30740	GmRFL22
Glyma09g30940	GmRFL23
Glyma09g39260	GmRFL24
Glyma09g39940	GmRFL25
Glyma10g00540	GmRFL26
Glyma14g38270	GmRFL28
Glyma15g24040	GmRFL29
Glyma16g25410	GmRFL30
Glyma16g27600	GmRFL31
Glyma16g27640	GmRFL32
Glyma16g27790	GmRFL33
Glyma16g27800	GmRFL34
Glyma16g28020	GmRFL35
Glyma16g31950	GmRFL36
Glyma16g31960	GmRFL37
Glyma16g32030	GmRFL38
Glyma16g32050	GmRFL39
Glyma16g32420	GmRFL40
Glyma18g46270	GmRFL41
jgi Poptr1 556096 eugene3.00040809 ^{*3}	PtRFL1
jgi Poptr1 561788 eugene3.00061747	PtRFL2
jgi Poptr1 561789 eugene3.00061748	PtRFL3
jgi Poptr1 562052 eugene3.00062011	PtRFL4
jgi Poptr1 570945 eugene3.00130309	PtRFL5
jgi Poptr1 570961 eugene3.00130325	PtRFL6
jgi Poptr1 570963 eugene3.00130327	PtRFL7
jgi Poptr1 571830 eugene3.00131194	PtRFL8
jgi Poptr1 573736 eugene3.00190210	PtRFL9
jgi Poptr1 581077 eugene3.01250075	PtRFL10
jgi Poptr1 581161 eugene3.127570001	PtRFL11
jgi Poptr1 594495 eugene3.00640083	PtRFL12
jgi Poptr1 595453 eugene3.00700094	PtRFL13
jgi Poptr1 595455 eugene3.00700096	PtRFL14
jgi Poptr1 595479 eugene3.00700120	PtRFL15
jgi Poptr1 595494 eugene3.00700135	PtRFL16
jgi Poptr1 595495 eugene3.00700136	PtRFL17
jgi Poptr1 595506 eugene3.00700147	PtRFL18
jgi Poptr1 562855 eugene3.00070793	PtRFL19
jgi Poptr1 572581 eugene3.00140626	PtRFL20
AAM52339 ^{*4}	PPR-592
CAD61285	radish-Rf
<i>bvORF16</i> ^{*5}	bvORF16

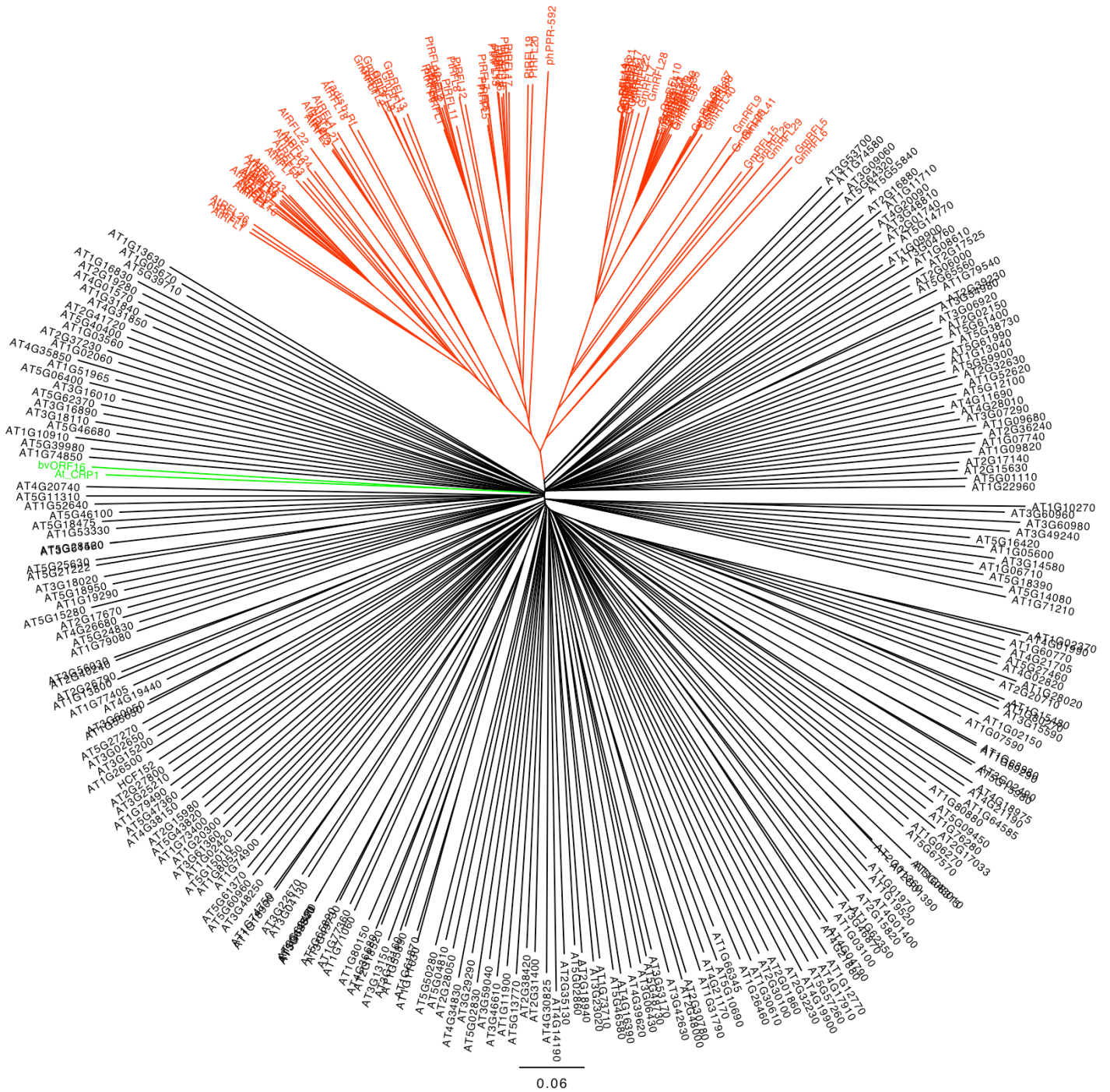
*1 <http://www.plantenergy.uwa.edu.au/applications/ppr/ppr.php> and <http://www.arabidopsis.org/tools/bulk/sequences/index.jsp>

*2 http://www.phytozome.net/search.php?show=text&org=Org_Gmax_v1.1

*3 http://www.phytozome.net/search.php?show=blast&method=Org_Ptrichocarpa_v2.2

*4 DDBJ/GenBank/EMBL dataase

*5 This study



File S2. Phylogenetic tree drawn by the Neighbor-Joining method. Amino acid sequences listed in File S1 were aligned using ClustalW (<http://clustalw.ddbj.nig.ac.jp/index.php?lang=ja>) and tree data were obtained. The tree was drawn using FigTree software (<http://tree.bio.ed.ac.uk/software/figtree/>). The tree includes: P-type PPR proteins from *Arabidopsis thaliana* (O'Toole et al., Mol. Biol. Evol., 2008, 25: 1120-1128); soybean PPR-type Rf-like (RFL) proteins (Fujii et al., PNAS, 2011, 108: 1723-1728); poplar RFL proteins (Fujii et al., PNAS, 2011, 108: 1723-1728); petunia RF protein (Bentolila et al., PNAS, 2002, 99: 10887-10892); radish RF protein (Brown et al. Plant J., 2003, 35: 262-272; Desloire et al., EMBO Rep., 2003, 4: 588-594; Koizuka et al., Plant J., 2003, 34: 407-415); and bvORF16. Clades including bvORF16 and AT5g42310 (At_CRP1), and RF and RFL are colored by green and red, respectively.


```

At_OMA1 MSWYRRTKLVFDSLRNINPKILPRSHVTSRINNPIGSSNPSAKFSSIS 50
Os_OMA1 MNYLKNSRSVLSRLLR-HKPTGCPRLPPSP----PLPQAPPAGYYFTSPS 45
BV_ORF19 MAWYRNSRFVYNALKLNLRSKTFGTIPTPR-----VHSNSSSLFYNQST 44
Sc_OMA1 M-----LRNIIRFKGFG-----KGTSGGFLKPVSF 25

At_OMA1 REVGLRSWTSLGRNTNRIAYNPFLSQPKRYYYVD--RYQVRHFKPRGPGR 98
Os_OMA1 RPEAVRFGRVLLRSP-PPPPRPAQAPPSRYFYTSPQRQKVVFNRRRGSR 94
BV_ORF19 NKCSGLFGSAKSGYFNGFKHHQEISSFSGFARRN-----YHGDKTEVSV 88
Sc_OMA1 R----VQLTRCYRYDNGPSYRRFNG-----EYSQKSSFK 56

At_OMA1 WFQNPRTVFTVVVLVGSVCLITLIVGNTETIPYTKRTHFILLSKPMEKLLG 148
Os_OMA1 WYHDPKRLTTVVVVVSGGAAAAYVFGNLETVPYTNRTHLILLSPPLERQLG 144
BV_ORF19 ESWLEKFLVPIGLILTFGILGYPHVHPVVVYPYTKRKHVVLMTTRENETG 138
Sc_OMA1 SILLDKSSRKYLALLFGCCSLFYTHLDAKAVSDRSRFTWVSRPLELTIC 106

At_OMA1 ETQFEQIKKTYQGKILPATHPESIRVRLTAKEVIDALORGLS----NERV 194
Os_OMA1 ESQFNNLKKELGPKILPLHPDSIRVRLTASEVVRVAVHRLAGRHDAFA 194
BV_ORF19 EVEKR-----KIQPATHPDTDRVRSIFOHILESLEREIN----- 172
Sc_OMA1 NYTYKSIWRQTQOEILPQHPLSIKIENIFMKIVEAAYKDPS----- 148

At_OMA1 WSDLGYASTESSLGGG-SDKGVKEMEMAMS--GEDTMTDMKWSKEDQVLD 241
Os_OMA1 ADDASYGDISTDVVIKNHEAGAEDVMLGRSRGNKNASVAAAQRDEEVL 244
BV_ORF19 -----HHELELELE---RDETFKEKTIWKEETVD 198
Sc_OMA1 -----

At_OMA1 DQWIQSRK--KDSKAHAATSHLEGISWEVLVVNEP IVN--AFCLPAGKI 287
Os_OMA1 DRWVTESRDRGKARGAQPETRHLDGLNWEVIVVRDDLIN--AMCLPGGKI 292
BV_ORF19 -----DKDSRKKHSGAKITINHLEGMNWEIFVVDKPLVE--SSYLLGGKI 241
Sc_OMA1 -----VDNSLIDGIKWEIHVVNDPTASPNAFVLPGGKV 181

At_OMA1 VVFTGLLNHFKSDAEVATVIGHEVGHAVARHVAEGITKNLWFATLQLV-I 336
Os_OMA1 VVFTGLLNHFKTDAEIATVIGHEVGHATARHAAEMI TKNLWFATLQIV-I 341
BV_ORF19 VVYTGLLNHCNSDAELATI IAHQVGHAVARHEAEDSTAFFWLLISLNVIL 291
Sc_OMA1 FIFSSITPICANDDGIATVLAHEFAHQARHTAENLSKAPIYSLIGLV-I 230

At_OMA1 -YQFV-MEDLVNTMSALFLRLPFPSR----- 359
Os_OMA1 -MQFIYMDMINAMSTLLKLPFSR----- 365
BV_ORF19 FKILFTPEESANARSKLLRHPLLOKVKI IQARAPQLLPTICLSLVGL 341
Sc_OMA1 -YTVTGAHAINNILLDGLRMPASR----- 254

At_OMA1 -----KMEIEADYIGLLLASAGYDPRVAPTVEKLG-----KLG 394
Os_OMA1 -----RMEIEADHIGLLVLGAAGYDPRVAPSVEKLG-----KIA 400
BV_ORF19 FSSVFILYYGRKEIEADHIGVLLMASAGYDPRVAPQVYDKLA-----KPL 386
Sc_OMA1 -----QMETEADYIGLMIMSRACFQEQESIKVWERMANFEKQMN 294

At_OMA1 GD-ALGDYLSSTHPSGKKRSLKLLAQANVMEALMIYREVQACRTGVEGFL-- 442
Os_OMA1 GDSTLSNYLSTHPSKKRAQLLRQAKVMDEALRLYREVSSGQ-GTEGFL-- 448
BV_ORF19 GD---WNCLATHPFARMRAKLLARADVMEADKIYNEVVAGR-AIQGLQ--- 431
Sc_OMA1 CGVVNMEFLSTHPASTRIENMSKWLPRANEIYEQSDCSSMGNYKSF 345

```

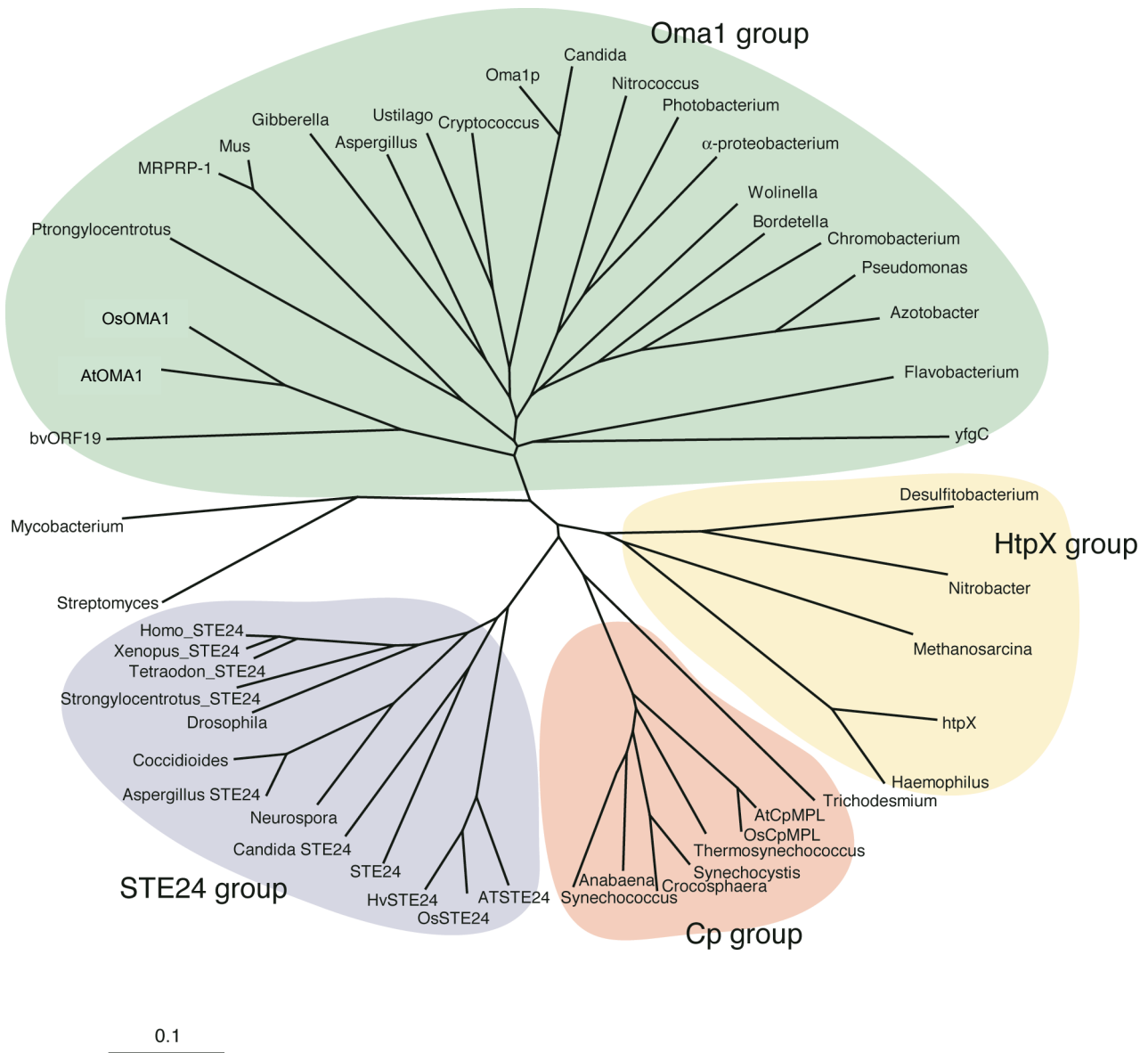
File S3. Multiple alignment of amino acid sequences of OMA1-homologous proteins from Arabidopsis (At_OMA1, At5g51740), rice (Os_OMA1, Os02g0735100), sugar beet (bvORF19, this study), and yeast (Sc_OMA1, S000001795). Position of the Zn²⁺ binding motif is shown by a horizontal line. The amino acid sequences were aligned using ClustalW (<http://clustalw.ddbj.nig.ac.jp/index.php?lang=ja>). The identity of amino acid sequences between bvORF19 and yeast OMA1 is 17%. The E-value obtained from a BLAST search using bvORF19 as a query is 1e-12 for yeast OMA1.

STE24
 Candida_STE24
 Aspergillus_STE24
 Coccidioides
 Neurospora
 OsSTE24
 HvSTE24
 ATSTE24
 Homo_STE24
 Mus_STE24
 Bos_STE24
 Gallus_STE24
 Xenopus_STE24
 Tetraodon_STE24
 Strongylocentrotus_STE24
 Drosophila
 AtCpMPL
 OsCpMPL
 Synechocystis
 Crocosphaera
 Anabaena
 Thermosynechococcus
 Synechococcus
 Streptomyces
 Mycobacterium
 Trichodesmium
 htpX
 Haemophilus
 Desulfitobacterium
 Nitrobacter
 Methanosarcina
 Wolinella
 Photobacterium
 a-proteobacterium
 Nitrococcus
 Pseudomonas
 Azotobacter
 Chromobacterium
 Bordetella
 Flavobacterium
 Ustilago
 Cryptococcus
 Omalp
 Candida
 MRPRP-1
 Mus
 Pstrongylocentrotus
 Gibberella
 Aspergillus
 AtMPL
 OsMPL
 ORF19
 yfgC

EI TAVLAHEI GHWQKNE IVNMVIFS QLHTFLIFSL
 ETVAVLAHEI GHWKLNE LPKMITMMQGHFLIFSL
 EVVAVLSHEL GHWSLGH TTKLFAIAQSHMFYIFAL
 EVVAVLSHEL GHWSLGH TTKLFGIAQFHMFYIFAL
 EVVAVLAHEL GHWKLGH TTSLFGISQAHFFAIFSL
 EIVSVIAHEL GHWKLNE TVYSFVAVQLLMFLQFGG
 EIVSVLAHEL GHWKLNE TAYSFVAVQLLTFMQFGG
 EIVAVIAHEL GHWKLNE TTYSFIAVQILAFQLQFGG
 EVLAVLGHEL GHWKLGH TVKNI IISQMNSFLCFFL
 EVLAVLGHEL GHWKLGH TVKNI IISQMNSFLCFFL
 EVLAVLGHEL GHWKLGH TVKNI IISQMNSFLCFFL
 EVLAVLGHEL GHWKLGH TVKNI IISQMNSFLCFFL
 EVLAVLGHEL GHWKLGH TVKNI IISQMNSFLCFFL
 EILAVLGHEL GHWKLGH TVKNI IISQMNSFLCFFL
 EVLAVLAHEL GHWKLGH NLKNI IISQVNL LCLFL
 EVLAVLGHEL GHWKLGH TVKNI IIMQVHLFLMFLV
 ELQAVLAHEL GH LKCDHGVWLT FANILT - - LGAYT
 ELQAVLAHEL GH LKCDHGVWLT FANILT - - MGAYS
 EI QAVMAHEL GH LKCEHGVYLT LANIMV - - LAAGL
 EI QGVMAHEL GH LKCEHGVYLT LANIMV - - LGASL
 EI QAVIAHEL GH LKCDHGVYLT PVNLLV - - LAASA
 ELQAVLAHEL GH LKCEHGVYLT IANLLL - - FAASQ
 EI QAVIAHEL GH LKCDHGVYLT MANLLM - - LSTSL
 EMRAVIGHEV GH ALSGH SVYRT ILLFLTSLALRVA
 EMRFVMGHEL GH ALSGH AVYRT MMMHLRLARSFG
 ELKTVLAHEL GH IKCGH PILLNQMA TWAMG IASAIT
 EAEAVIAHEI SH IANGDMVTMT LIQGVVNTFVIFI
 EAEAVLAHEI SH ISNGDMVTMALLQGV LNTFVIFL
 ELEGVLAHEI MAHIKNRD ILI STLAA - VMAGVI TTL
 ELAGVIAHEI LAHIK HHD TLLMTITA - T IAGAI SML
 ELEAVLAHEI SHVKNRD MAVLT IAS - FLS SVAFYI
 ELAVVMGHEI AHAIARE GAERLSVSMA SELGRNLI
 QLAVIGHEI GHVIAQH SNERLSRSQLANAGLELT
 QLAVVMGHEI GHVIAEH GNERMSIATL SNLGLQIT
 QLAVIGHEV GHVLA GH ANERL STNAA TQTGLDLL
 EIAAVMGHEI AHALREH GREAMSKAYGVQVASQ - I
 EIAAVMGHEI AHALREH GREALSKAYAVEMAKQGA
 ELAAVIGHEI SHALREH TRENMSQAYAQMGGLGLV
 ELAAVIGHEI AHALREH ARERV SQQMATSI GLSVL
 GLAMILGHEI AHALANE GAQRMTAQGGQ I V G A G
 GLATVIGHEI AHQVARE SA EKMSGYK VLLFGTFL
 GLATVIGHEI AHQVARE PAERMSSMKVLFALG LLL
 GIATVLAHEI FAHQVARE TAENLSKAPIYSLGLGLV
 GIATVLSHEI FAHQVARE TAENLSKAPLYSLGLIIL
 QLSFLLGHEI AHAVLGH AA EKAGMVHLDFLGMIF
 QLSFLLGHEI AHAVLGH AA EKASLVHLDFLGMIF
 QLGTVLAHEI MAHVVLNE SAEMASFFEFDFL FMIVV
 ALAAVIGHEI AHNTA SHASERL SA AWVGNLTA GSL
 GLAAVIGHEI AHVVAHEI TGERMSN - - - NFVTMGV
 EVATVIGHEV GHAVARE VA EGI TKNLWFAI - LQLV
 EIAVIGHEV GHAIARE AAEMI TKNLWFWI - LQIV
 ELATIIAHQV GHAVARE EAEDSTA FFWLLI SLNVI
 QLAVMAHEI SHVTQREI LARAMEDQQR SAPLTVWG

File S4. See next page for the legend.

File S4. Multiple alignment of ~35 amino acid residues surrounding the Zn²⁺ binding motif of peptidase M48 proteins, a protein family to which yeast OMA1 belongs. The position of the Zn²⁺ binding motif is shown by a horizontal line. Note that only bvORF19 (indicated by ORF19 in the alignment) contains HQxxH, instead of HEXxH that is present in the other members. Data from: STE24, *Saccharomyces cerevisiae*, CAA89647; *Candida*_STE24, *Candida albicans*, XP_713382; *Aspergillus*_STE24, *Aspergillus fumigatus*, XP_752066; *Coccidioides*, *Coccidioides immitis*, EAS28348; *Neurospora*, *Neurospora crassa*, CAC28689; OsSTE24, *Oryza sativa*, Os02g0680400; HvSTE24, *Hordeum vulgare*, CAL26913; ATSTE24, *Arabidopsis thaliana*, At4g01320; Homo_STE24, *Homo sapiens*, NP_005848; Mus_STE24, *Mus musculus*, NP_766288; Bos_STE24, *Bos taurus*, XP_882083; Gallus_STE24, *Gallus gallus*, XP_417720; Xenopus_STE24, *Xenopus laevis*, AAH82484; Tetradon_STE24, *Tetraodon nigroviridis*, CAG10466; *Strongylocentrotus*_STE24, *Strongylocentrotus purpuratus*, XP_001177479; *Drosophila*_STE24, *Drosophila melanogaster*; AtCpMPL, *Arabidopsis thaliana*, At3g27110; OsCpMPL, *Oryza sativa*, Os01g0970700; *Synechocystis*, *Synechocystis* sp. PCC 6803, NP_440889, *Crocospaera*, *Crocospaera watsonii*, NP_681428; *Anabaena*, *Anabaena variabilis*, YP_321952; *Thermosynechococcus*, *Thermosynechococcus elongatus*, NP_681428; *Synechococcus*, *Synechococcus* sp. JA-3-3Ab, YP_473883; *Streptomyces*, *Streptomyces avermitilis*, NP_826653; *Mycobacterium*, *Mycobacterium tuberculosis*, NP_216493; *Trichodesmium*, *Trichodesmium erythraeum*, YP_721635; htpX, *Escherichia coli*, AAA62779; *Haemophilus*, *Haemophilus influenzae*, NP_438878; *Desulfitobacterium*, *Desulfitobacterium hafniense*, ZP_01369144; *Nitrobacter*, *Nitrobacter hamburgensis*, YP_575597; *Methanosarcina*, *Methanosarcina mazei*, NP_635158; yfgC, *Escherichia coli*, AAC75547; *Desulfovibrio*, *Desulfovibrio desulfuricans*, YP_386603; *Wolinella*, *Wolinella succinogenes*, NP_907498; *Photobacterium*, *Photobacterium profundum*, YP_132334; a-proteobacterium, a-proteobacterium HTCC2255, ZP_01448796; *Nitrococcus*, *Nitrococcus mobilis*, ZP_01126393; *Pseudomonas*, *Pseudomonas aeruginosa*, NP_253322; *Azotobacter*, *Azotobacter vinelandii*, ZP_00416091; *Chromobacterium*, *Chromobacterium violaceum*, NP_899823; *Bordetella*, *Bordetella bronchiseptica*, NP_888655; *Flavobacterium*, *Flavobacterium* sp. MED217, ZP_01061128; *Ustilago*, *Ustilago maydis*, XP_757961; *Cryptococcus*, *Cryptococcus neoformans*, XP_569916; Oma1p, *Saccharomyces cerevisiae*, P36163; *Candida*, *Candida glabrata*, XP_446463; MPRP-1, *Homo sapiens*, BAC79381; *Mus*, *Mus musculus*, NP_080185; *Strongylocentrotus*, *Strongylocentrotus purpuratus*, XP_799173; *Gibberella*, *Gibberella zeae*, XP_390368; *Aspergillus*, *Aspergillus nidulans*, XP_659454; AtMPL, *Arabidopsis thaliana*, At5g51740 (AtOMA1); OsMPL, *Oryza sativa*, Os02g0735100 (OsOMA1). Multiple alignment was done by using ClustalX (<http://www.clustal.org/clustal2/>).



File S5. A Neighbor-Joining tree of peptidase M48 family proteins (see File S4). The tree was drawn by TreeView (<http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>) based on the alignment shown in File S4. The sequence data are grouped into four clades. Note that *bvORF19*, as well as its homologous sequences in *Arabidopsis*, rice, and yeast *OMA1* (see File S3), belongs to a single group, tentatively named the Oma1 group.