### Unusual and Typical Features of a Novel Restorer-of-Fertility Gene of Sugar Beet (Beta vulgaris L.)

Hiroaki Matsuhira,\*<sup>1</sup> Hiroyo Kagami,\* Masayuki Kurata,\* Kazuyoshi Kitazaki,\*<sup>2</sup> Muneyuki Matsunaga,\* Yuko Hamaguchi,\* Eiki Hagihara,\* Minoru Ueda,\* Michiyo Harada,\* Aki Muramatsu,\* Rika Yui-Kurino,\* Kazunori Taguchi,<sup>†</sup> Hideto Tamagake,<sup>‡</sup> Tetsuo Mikami,\* and Tomohiko Kubo\*<sup>3</sup>

\*Laboratory of Genetic Engineering, Research Faculty of Agriculture, Hokkaido University, Sapporo, Hokkaido 060-8589, Japan, <sup>†</sup>Memuro Upland Farming Research Station, Hokkaido Agricultural Research Center, National Agriculture and Food Research Organization, Memuro, Hokkaido 082-0081, Japan, and <sup>‡</sup>Central Agricultural Experiment Station, Agriculture Research Department, Hokkaido Research Organization, Takikawa 073-0013, Japan

**ABSTRACT** Male gametogenesis in plants can be impaired by an incompatibility between nuclear and mitochondrial genomes, termed cytoplasmic male sterility (CMS). A sterilizing factor resides in mitochondria, whereas a nuclear factor, *Restorer-of-fertility* (*Rf*), restores male fertility. Although a majority of plant *Rf* genes are thought to encode a family of RNA-binding proteins called pentatrico-peptide repeat (PPR) proteins, we isolated a novel type of *Rf* from sugar beet. Two BACs and one cosmid clone that constituted a 383-kbp contig covering the sugar beet *Rf1* locus were sequenced. Of 41 genes borne by the contig, quadruplicated genes were found to be associated with specific transcripts in *Rf1* flower buds. The quadruplicated genes encoded a protein resembling OMA1, a protein known from yeast and mammals to be involved in mitochondrial protein quality control. Construction of transgenic plants revealed that one of the four genes (*bvORF20*) was capable of restoring partial pollen fertility to CMS sugar beet; the level of restoration was comparable to that evaluated by a crossing experiment. However, the other genes lacked such a capability. A GFP-fusion experiment showed that *bvORF20* encoded a mitochondrial protein. The corresponding gene was cloned from *rf1rf1* sugar beet *Rf1*, such as gene clustering and copy-number variation between *Rf1* and *rf*, were reminiscent of PPR-type *Rf*, suggesting that a common evolutionary mechanism(s) operates on plant *Rfs* irrespective of the translation product.

As a phenotypic manifestation of nuclear-mitochondrial incompatibility in plants, cytoplasmic male sterility (CMS) has garnered much interest and has been recorded to occur in >140 plant species (Laser and Lersten 1972). CMS is a maternally inherited trait that inactivates male reproductive function in otherwise normal plants (Schnable and Wise 1998). A genetic model developed to explain CMS suggests that it involves a nuclear–mitochondrial interaction in which a sterility-inducing factor (S) is generated in mitochondria, and one or more nuclear factors, termed *restorers of fertility* (*Rf*), capable of inhibiting the action of S (Hanson and Bentolila 2004). According to this model, plants with the S factor and two nonrestoring nuclear alleles, *i.e.*, [S] *rfrf*, are male sterile (MS), whereas [S]*RfRf* or [S]*Rfrf* plants produce functional pollen (Budar *et al.* 2006; Chase 2007). Plants with N mitochondria lack the S factor and are male fertile irrespective of their nuclear alleles in the *Rf* locus.

Many S factors have been associated with various unique polypeptides encoded by mitochondrial genomes (Pelletier and Budar 2007). In some cases, the evolutionary origin of the S factor is unclear because the mitochondrial ORF that encodes the unique polypeptide (S-ORF) has no homology within the N mitochondrial genome or with any nucleotide

Copyright © 2012 by the Genetics Society of America

doi: 10.1534/genetics.112.145409

Manuscript received June 25, 2012; accepted for publication September 10, 2012 Available freely online through the author-supported open access option. Supporting information is available online at http://www.genetics.org/lookup/suppl/

doi:10.1534/genetics.112.145409/-/DC1. <sup>1</sup>Present address: Crop Breeding Research Division, NARO-HARC, Sapporo,

Hokkaido 062-8555, Japan.

<sup>&</sup>lt;sup>2</sup>Present address: Central Research Institute of Electric Power Industry, Abiko, Chiba 270-1194, Japan.

<sup>&</sup>lt;sup>3</sup>Corresponding author: Laboratory of Genetic Engineering, Research Faculty of

Agriculture, Hokkaido University, N-9, W-9, Kita-ku, Sapporo, Hokkaido 060-8589, Japan. E-mail: gelab@abs.agr.hokudai.ac.jp.

sequences known to date. In other cases, S-ORFs appear to be mosaic of parts of duplicated mitochondrial genes, suggesting that S-ORFs are by-products of mitochondrial genome rearrangement (for reviews, Budar *et al.* 2004; Kubo and Newton 2008; Kubo *et al.* 2011).

Nuclear Rfs seem to overcome the action of S factors in different ways, but the mechanisms are obscure. One group of Rfs regulates the expression of S-ORFs at the posttranscriptional level (Fujii and Toriyama 2008). Plants having this type of Rf accumulated fewer S-ORF polypeptides with or without an altered level of S-ORF transcription. Molecular cloning of such Rfs from petunia (Petunia X hybrida hort. ex Vilm), radish (Raphanus sativus L.), and rice (Oryza sativa L.) revealed that these genes encode a class of proteins sharing a common sequence termed a pentatricopeptide repeat (PPR) (Bentolila et al. 2002; Brown et al. 2003; Desloire et al. 2003; Kazama and Torivama 2003; Koizuka et al. 2003; Akagi et al. 2004; Komori et al. 2004; Wang et al. 2006; Hu et al. 2012). These proteins constitute a large gene family that is associated with post-transcriptional gene regulation in plant organelles (Schmitz-Linneweber and Small 2008). A genetic association of *Rf* loci with PPR genes also has been reported from other plants such as CMS-S maize (Zea mays L.), sorghum [Sorghum bicolor (L.) Moench] and Mimulus (Klein et al. 2005; Xu et al. 2009; Barr and Fishman 2010; Jordan et al. 2010).

Rfs distinct from the PPR type, are known, but the current paucity of knowledge precludes further classification. Three non-PPR-type Rfs have been identified to date: maize Rf2a, rice Rf17, and rice Rf2. Maize Rf2a was the first Rf cloned, and encodes a mitochondrial aldehyde dehydrogenase (Cui et al. 1996). However, the functional relationship between URF13-T, a polypeptide encoded by the S-ORF in maize T-type CMS (Dewey et al. 1986), and RF2A proteins is unclear. Rice Rf17 was cloned as an Rf for CW-type CMS (Fujii and Toriyama 2009). The reduced expression of Rf17 in CW mitochondria compromises MS expression, thereby functionally acting as if *Rf17* restored male fertility. It remains unknown whether any direct relationships exist between Rf17 and an, as yet, unidentified S-ORF in CW-CMS mitochondria. Genes for glycine-rich proteins have been isolated as rice Rf2 for Lead Rice (LD)-type CMS via map-based cloning (Itabashi et al. 2011). Hu et al. (2012) reported that a PPR-type RF protein, a glycine-rich protein, and a transcript encoding S factor are components of a large mitochondrial complex of 400-500 kDa in Hong-Lian (HL)-type CMS in rice.

Given its importance in hybrid seed production, sugar beet CMS has been extensively studied (Boutry *et al.* 1984; Lind *et al.* 1991; Hallden *et al.* 1992; Ducos *et al.* 2001). CMS mitochondria of sugar beet are characterized by a unique 39-kDa polypeptide encoded by an N-terminal extension of *atp6* (*preSatp6*) that is missing in N mitochondria (Yamamoto *et al.* 2005). A precursor polypeptide consisting of preSATP6 and ATP6 is hypothesized to be cleaved into two separate polypeptides, one being the mature ATP6 polypeptide, and the other a preSATP6 polypeptide which subsequently forms a 200-kDa oligomer in the mitochondrial membrane. However, following fertility restoration, the amount of the preSATP6 polypeptide remained unchanged (Yamamoto *et al.* 2005), an observation that led us to postulate the involvement of a non–PPR-type *Rf*.

According to a genetic model proposed by Owen (1945), fertility restoration in sugar beet requires two independent genes, X and Z, of which the latter seemed less effective. Genetic mapping of X and Z located these genes on chromosomes III and IV, respectively (Pillen et al. 1993; Schondelmaier and Jung 1997; Hjerdin-Panagopoulos et al. 2002; Bosemark 2006). We previously found that pollen fertility segregated as if it were controlled by a single dominant gene when the sugar beet line NK-198 was used as a pollen parent (Hagihara et al. 2005a), although the level of fertility restoration varied depending on the nuclear genetic background (Hagihara et al. 2005a). The NK-198 Rf was named Rf1 and mapped to a terminal region of chromosome III, suggesting that the Rf1 was an allele of the X locus (Hagihara et al. 2005a). Molecular markers linked to Rf1 were used to isolate BAC clones that covered the Rf1 locus (Hagihara et al. 2005b).

In this study, the nucleotide sequence of a 383-kbp chromosomal region containing the sugar beet *Rf1* was determined. From this sequence, we found that an unexpected gene satisfied the following criteria: specific transcription in *Rf1* flower buds, partial fertility restoration to transgenic sugar beet (the level of restoration is comparable to that evaluated by a crossing experiment), and mitochondrial localization of the GFP-fused protein. The gene was related to yeast *Oma1* known to be involved in quality control of mitochondrial proteins (Kaser *et al.* 2003). We also found an organizational similarity between sugar beet *Rf1* locus and some PPR-type *Rf* loci in terms of gene clustering and copynumber variation between *Rf1* and *rf1*, suggesting that a common evolutionary mechanism(s) operates on plant *Rfs*.

### **Materials and Methods**

### Plant materials

A restorer line NK–198, three maintainer lines TK–81mm–O, TA–33–O, and NK–219mm–O, and a CMS line NK–219mm–CMS used in this study were developed at the Hokkaido Agricultural Research Center, National Agriculture and Food Research Organization, Japan. Crosses were made by exchanging paper bags over the inflorescences in a greenhouse. Plants were vernalized for 4 months (5°, 24 hr/day) and flowered in the greenhouse. Anther tissues were sampled to examine pollen fertility on the day of anthesis. Pollen fertility was examined by Alexander staining (Alexander 1969).

### Isolation of nucleic acids

Total cellular DNA of beet plants was isolated from fresh green leaves by the CTAB-based method described by Doyle and Doyle (1990). DNAs from BAC clones, cosmid clones, and plasmid clones were isolated by an alkaline lysis procedure (Sambrook *et al.* 1989). Lambda-phage DNA was

isolated by a liquid culture method (Sambrook *et al.* 1989). Isolated DNA was purified by cesium chloride-ethidium bromide (CsCl-EtBr) equilibrium centrifugation when necessary. Total RNA from sugar beets was isolated according to Chomczynski and Sacchi (1987) or by using the RNeasy Plant Mini kit (Qiagen, Hilden, Germany). Residual DNA in the RNA sample was removed by DNase I (Takara Bio, Ohtsu, Japan) digestion in the presence of 8 mM MgCl<sub>2</sub>.

### Subcloning into a cosmid vector

Purified BAC-clone DNA was partially digested with Sau3A I (Takara Bio), then electrophoresed in an agarose gel. DNA fragments of 30-50 kbp were eluted from the gel and partially filled to obtain a 5'-GA-3' end (0.5 M Tris-HCl pH 7.5, 100 mM MgCl<sub>2</sub>, 10 mM dithiothreitol, 80 μM dATP, 80 μM dGTP, 2 units Klenow fragments, 30 min at room temperature) to prevent self-ligation. The cosmid vector pWE15 (Stratagene, La Jolla, CA) was completely digested with XhoI and then partially filled to obtain a 5'-TC-3' end (0.5 M Tris-HCl pH 7.5, 100 mM MgCl<sub>2</sub>, 10 mM dithiothreitol, 80 µM dCTP, 80 µM dTTP, 2 units Klenow fragments, 30 min at room temperature) to prevent self-ligation. The ligation reaction was carried out using T4 DNA ligase (New England Biolabs, Beverly, MA) in the presence of 10% polyethylene glycol 8000. The ligated DNA sample was precipitated with ethanol and then dissolved in water. Gigapack III Gold (Stratagene) was used for packaging.

# Construction of the shotgun library and nucleotide sequencing

Inserts of the lambda-phage clone were amplified with LA-Taq (Takara Bio) according to the instruction manual. Inserts of the cosmid clone were cut out by *Not*I digestion and recovered from gel slices after electrophoresis. The inserts or whole BAC-clone DNAs were randomly sheared by sonication and then electrophoresed in an agarose gel. DNA fragments of 1.2–1.5 and 2.0–2.5 kbp were eluted from the gel slices. The ends of DNA fragments were blunted by T4 DNA polymerase (Takara Bio) in the presence of dATP, dTTP, dCTP, and dGTP, and then ligated into the *Hinc*II site of pUC19. Plasmid DNA was sequenced using a LIC-4200L (Li-COR, Lincoln, NE) or ABI3130 (Applied Biosystems, Foster City, CA) sequencer.

### **Bioinformatics**

Assembly of the nucleotide sequence was done using a Staden package (Staden 1996) and Sequencher 4.0 (Hitachi Software Engineering, Tokyo). Protein-coding regions were predicted by GENESCAN (Burge and Karlin 1997) (http://genes.mit. edu/GENSCAN.html) with an *Arabidopsis* matrix and the BLASTX program (http://www.ncbi.nlm.nih.gov/). A homology search for putative amino acid sequences was done using BLASTP on the National Center for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov/). Intracellular localizations were predicted using TargetP (Emanuelsson *et al.* 2000) (http://www.cbs.dtu.dk/services/TargetP/) and Predotar (Small *et al.* 2004) (http://urgi.versailles.inra.fr/predotar/

predotar.html). A motif search was undertaken using Pfam (Finn *et al.* 2006) (http://pfam.sanger.ac.uk/). Repeated sequences were searched with Reputer (Kurtz *et al.* 2001) (http://bibiserv.techfak.uni-bielefeld.de/reputer/). Multiple sequences were aligned using ClustalW (Chenna *et al.* 2003) (http://clustalw.ddbj.nig.ac.jp/top-j.html). Nucleotide sequences reported in this study are deposited in the DNA Data Bank of Japan (DDBJ)/GenBank/EMBL under accession nos. AB646133 (4F1), AB646134 (5A3), AB646135 (33E19), and AB646136 (no. 10).

### PCR and direct sequencing

Total cellular DNA (5–10 ng) was subjected to PCR amplification using LA-Taq (Takara Bio) or GoTaq Green Master mix (Promega, Madison, WI). Total RNA (2  $\mu$ g) was reverse transcribed with the SuperScript III First-Strand Synthesis system (Invitrogen, Carlsbad, CA). The resultant cDNA was subjected to PCR amplification. Direct sequencing was achieved using an ABI3130 sequencer (Applied Biosystems).

### Hybridization

Colony- and plaque-lift filters were prepared using Hybond N+ membranes (GE Healthcare, Amersham Place, UK) according to the instruction manual. For DNA gel blot analysis, a DNA sample (5 µg) was digested with restriction endonucleases purchased from Takara Bio and electrophoresed in a 1% agarose gel. After denaturation and neutralization, DNA fragments were transferred to Hybond N+ membranes according to the instruction manual. For RNA gel blot analysis, 5  $\mu$ g RNA was electrophoresed in a 1.5% agarose gel containing 0.66 M formaldehyde and then transferred by capillary action to Hybond N+. The DNA fragment of interest was labeled with <sup>32</sup>P using the Megaprime DNA labeling system (GE Healthcare) or with alkaline phosphatase using the AlkPhos Direct DNA labeling system (GE Healthcare). Hybridization was conducted according to the manufacturer's instructions. Signal bands were detected on X-ray films or with an image analyzer (BAS2000; Fuji Photo Film, Tokyo).

### Construction of GFP-fusion genes and transient assays

The pTH2 cloning vector, whose NcoI site includes the initiation codon for GFP, was used (Chiu et al. 1996). Gene segments of interest were PCR amplified with a set of primers, one bearing a SalI and the other an NcoI target sequence (see Supporting Information, Table S1) so that the amplified ORF could fuse in-frame with GFP. The resultant PCR fragments were digested with SalI and NcoI and then ligated into pTH2. A fluorescent signal in mitochondria resulted from the expression of an Arabidopsis F1-ATPase δ-subunit-RFP fusion protein expressed from pMt-R, a derivative plasmid of pWs (Arimura and Tsutsumi 2002). A PCR fragment corresponding to the first 58 amino acids of Arabidopsis RuBisCo activase was amplified and then substituted for the Arabidopsis F1-ATPase  $\delta$ -subunit region of pMt-R. The resulting plasmid was designated pCp-R (Kitazaki et al. 2011). Plasmid DNA was ethanol precipitated with gold particles of 1 µm diameter (Bio-Rad Laboratories) and then introduced into the epidermal cells of onion bulbs or Welsh onion sheaths using a GIE-III IDERA system (Tanaka, Ishikari, Japan). The fluorescent signal was captured with a BX50 microscope system combined with a digital camera (DP70; Olympus, Tokyo).

### Generation of transgenic sugar beets

Genomic DNA fragments containing bvORF19, bvORF20, and bvORF21 were PCR amplified from BAC clone 9C23 (see Table S1 for primer information). Using BP Clonase Enzyme mix (Invitrogen), the genomic DNA fragments were cloned into the donor vector, pDONRzeo, according to the manufacturer's instruction manual. After verifying the sequence integrity, the inserted DNA fragments were transferred to the binary vector, pMDC123, encoding the bialaphos-resistance gene as a selectable marker (Curtis and Grossniklaus 2003) by using LR Clonase Enzyme mix (Invitrogen). A 5.3-kbp BglII fragment containing bvORF18 was obtained from cosmid clone 4F1 and subcloned into the BamHI site of pBluescript. After verifying the nucleotide sequence, the fragment was excised as a PstI-XbaI fragment and cloned into pMDC123. All the constructs were introduced into Agrobacterium tumefaciens strain LBA4404.

The generation of transgenic sugar beets was accomplished according to an unpublished procedure developed by H. Tamagake (unpublished data). Briefly, leaf explants from aseptic plantlets were laid onto a callus-inducing medium (based on the modified MS medium, where NH<sub>4</sub>NO<sub>3</sub> and 2-(morpholin-4-yl)ethanesulfonic acid (MES) were adjusted 825.0 mg/liter and 250 mg/liter, respectively), containing 0.25 mg/liter 6-benzyladenine (BA) and 2.5 g/liter gellan gum. White, friable calli were cultured in a suspension medium (the modified MS medium containing 0.25 mg/liter BA) for 10 days. After that, calli were co-cultured with Agrobacterium in the suspension medium containing 100 mg/liter acetosyringone for 3-4 days. The calli were washed with the suspension medium containing 100 mg/liter meropenem and 2 mg/liter bialaphos and transferred onto a selection medium (the modified MS medium containing 0.25 mg/liter BA, 8.0 g/liter agar, 50 mg/liter meropenem and 100 mg/ liter bialaphos). Calli resistant to bialaphos were regenerated into plantlets on a regeneration medium (the modified MS medium containing 1.0 mg/liter BA, 1.0 mg/liter 2,3,5triiodobenzonic acid, 1.0 mg/liter abscisic acid, 8.0 g/liter agar, 50 mg/liter meropenem, and 2 mg/liter bialaphos).

### Nucleotide sequences of oligonucleotides

Oligonucleotides used in this study are listed in Table S1 and Figure S1.

### Results

### Nucleotide sequence of the chromosomal region containing sugar beet Rf1

The sugar beet *Rf1* had previously been located to a region delimited by two molecular markers, mP-A16 and mCP-L6

(Hagihara *et al.* 2005b). The region was covered with an array of ordered BAC clones (Hagihara *et al.* 2005b). To obtain a nucleotide sequence of this region, we selected three of the clones, 5A3, 9C23, and 33E19 (Hagihara *et al.* 2005b), as sequencing templates. To minimize sequence redundancy, we screened a cosmid clone bridging 5A3 and 33E19, from a sublibrary made from 9C23 by using probes made up of 5A3- and 33E19-BAC ends. As a result, cosmid clone 4F1 was selected for sequencing.

We conducted shotgun sequencing of 5A3, 33E19, and 4F1, yielding 3047, 8058, and 164 independent plasmid sequences, respectively. The plasmid sequences were assembled into three sequences of 156,315; 201,705; and 36,977 bp, respectively. The average coverage was 10.79 for 5A3, 22.64 for 33E19, and 6.2 for 4F1. Overlaps of 4091 bp and 7539 bp occurred between the 4F1 and 5A3 sequences and between the 4F1 and 33E19 sequences, respectively. Therefore, the assembly of 5A3, 33E19, and 4F1 provided a continuous 383,367-bp sequence, with a G + C content of 34.9%. Sequence analysis revealed that target sequences of the five molecular markers (mCP-A54, mP-A16, mCP-K2, mCP-L6, and mCP-L45) that had been mapped to the vicinity of *Rf1* (Hagihara *et al.* 2005b) were included in the assembly in the order predicted by genetic analysis (Figure 1).

# Potential protein-coding genes in the sequenced region and their transcription

Sequence analysis of the 383,367-bp region identified three potential transposable elements (TEs) (Figure 1). One TE was homologous to the maize mutator element and its related TEs, and was named bvMULE-1 (Beta vulgaris Mutatorlike element) (Figure S2). The second TE contained two ORFs (Figure S3). The upstream ORF that encoded 752 amino acid residues (ORF-A in Figure S3) had no homology to any entries in public databases, but a Pfam search identified an RNA recognition motif (RRM). The putative translation product of the second ORF (ORF-B, 1297 amino acid residues) had a high homology to reverse transcriptases of plant long interspersed nuclear elements (LINEs), which include an endonuclease/exonuclease/phosphatase family domain and an RNA-dependent DNA polymerase domain. This structure resembles a group of sugar beet LINEs called BNR (Heitkam and Schmidt 2009). The third TE contained a 4701-bp ORF exhibiting a high homology to Ty3-gypsy-type retroelements, and was named bvgypsy-1 (Figure S4).

Aside from the ORFs encoded by the TEs, 41 genes were predicted. These were named *bvORF1–bvORF41* (Figure 1 and Table 1). We surveyed the rest of the sequenced region by BLASTX search to detect any homologous entries in the DDBJ/EMBL/GenBank database but found none. To infer the function of the 41 genes, we conducted a BLASTP search against the DDBJ/EMBL/GenBank database using each of their putative translation products as queries. Although 34 queries matched well with known plant proteins, 7 had no homology to any entries (Table 1). We obtained little information on the possible functions of 3 of the 34 queries



**Figure 1** Organization of a 383-kbp chromosomal region of NK–198 deduced from two BAC clones and a cosmid clone. *Hind*III restriction sites are shown as triangles. Horizontal arrows indicate predicted genes and their orientation; intronic sequences are omitted. Gray arrows denote the absence of any homologous genes in the database, whereas blue and green arrows indicate the presence of homologous genes in other plants with or without functional assignment, respectively. Orange boxes represent transposable elements, and red boxes show their neighboring repeated sequences. Positions of five molecular markers that were described in Hagihara *et al.* (2005b) are indicated by vertical arrows.

with known homologs, as no detailed studies of their homologous entries have been published. The remaining 31 queries retrieved homologous entries whose functions have been fairly well described. Of these entries, Table 1 lists the best matching putative function from the *Arabidopsis* genome entries and their description from the The Arabidopsis Information Resource (TAIR) database (http://www. arabidopsis.org/).

Because *Rf1* is a gene for male-fertility restoration, expression patterns of these genes in anthers helps narrow down the coding region of Rf1. RNA samples from NK-198 anthers, leaves, and roots were subjected to reverse transcription (RT)-PCR analysis. Primers for bvORF12-bvORF35, genes located in the region delimited by genetic markers mP-A16 and mCP-L6 (Hagihara et al. 2005b) (see Figure 1), were designed: a single primer set was expected to amplify *bvORF18–bvORF21* because these genes were very similar (Figure S5) (quadruplicated genes). Results of the 21 RT-PCR analyses are summarized in Table 1 (see also Figure S1). Transcripts of all genes except bvORF22, bvORF28, bvORF29, and bvORF34 were detected in all organs examined. No amplicon was observed in any organs when the bvORF28- or the bvORF29-specific primer set was used, whereas organ-specific expression was observed in bvORF22 and bvORF34, whose transcript levels were below the detection limit in leaves and roots, respectively.

There was a PPR protein gene in the 383-kbp region. Transcripts of this gene, *bvORF16*, were detected in NK–198 anthers (Table 1). However, because of the amino acid sequence homology between bvORF16 and at5g42310 (Table 1), which presumably is an ortholog of maize *crp1*, a regulatory gene of plastids (Barkan et al. 1994; Schmitz-Linneweber et al. 2005; Williams-Carrier et al. 2008), it seemed likely that bvORF16 encodes a plastid protein and not a mitochondrial protein. Two programs, TargetP and Predotar, predicted no specific localization for the bvORF16 translation product. We constructed a chimeric GFP gene with 80 N-terminal amino acid residues of bvORF16. The chimeric GFP genes were placed under the control of the 35S promoter of the cauliflower mosaic virus. We bombarded epidermal cells of Welsh onion sheath with plasmids carrying the chimeric GFP gene and observed fluorescent signals. Surprisingly, each of the localized green signals matched with either mitochondria or plastids that were marked by a mitochondrion-targeting RFP or a plastid-targetingRFP (see Materials and Methods), respectively (Figure 2, A–F). Therefore, bvORF16 encodes a dual-targeted PPR protein. As far as we know, no PPR-type Rf reported to date has exhibited this dual-targeting property (Bentolila et al. 2002; Wang et al. 2006). PPR-type Rfs and PPR-type Rflike (RFL) genes tend to cluster with similar genes on chromosomes (Fujii et al. 2011), unlike bvORF16, a single copy gene in the sugar beet genome (Figure S6). The PPRtype Rfs identified to date belong to a subclass of PPR genes (termed P class) and form a single clade with RFL genes in a phylogenetic tree of P-class PPR genes (Fujii et al. 2011). We examined whether bvORF16, which appears to be a P-class PPR gene, belongs to the clade of

#### Table 1 Characteristics of the genes identified in the 383-kbp region

		Best matched Arabidopsis entries		Т	ranscripts <sup>a</sup>	
Name of ORFs	Locus name	Description <sup>6</sup>	E-value	Anthers	Leaves	Roots
bvORF1	At2g04940	Scramblase related	<i>e</i> -80	NDC	ND	ND
bvORF2	At4g33260	Putative cdc20 protein	0	ND	ND	ND
bvORF3	At5g17210	Unknown function	5e-44	ND	ND	ND
bvORF4	NA <sup>d</sup>	No hit	NA	ND	ND	ND
bvORF5	At5g57020	N-myristoyltransferase	0	ND	ND	ND
bvORF6	NA	No hit	NA	ND	ND	ND
bvORF7	At5g17170	Enhancer of sos3-1 (ENH1)	4e-26	ND	ND	ND
bvORF8	At4g19490	Putative homolog of yeast Vps54	e-139	ND	ND	ND
bvORF9	At4g19490	Putative homolog of yeast Vps54	e-64	ND	ND	ND
bvORF10	At3g10520	Class 2 nonsymbiotic hemoglobin	2e-63	ND	ND	ND
bvORF11	At2g34780	MEE22, EMB1611, etc.	4e-79	ND	ND	ND
bvORF12	At1g65810	P loop containing nucleoside triphosphate hydrolases superfamily protein	0	+	+	+
bvORF13	At1g65810	P loop containing nucleoside triphosphate hydrolases superfamily protein	0	+ <sup>e</sup>	+	+
bvORF14	NA	No hit	NA	+	+	+
bvORF15	At3g03150	Unknown function	3e-15	+	+	+
bvORF16	At5g42310	Pentatricopeptide repeat (PPR-like) superfamily protein	4e-94	+	+	+
bvORF17	At3g49010	60S ribosomal protein L13	8e-82	+	+	+
bvORF18	At5g51740	Peptidase M48 family protein	6e-62	+	+	+
bvORF19	At5g51740	Peptidase M48 family protein	4 <i>E</i> -52	+	+	+
bvORF20	At5g51740	Peptidase M48 family protein	8 <i>E</i> -61	+	+	+
bvORF21	At5g51740	Peptidase M48 family protein	6 <i>E</i> -62	+	+	+
bvORF22	At3g50170	Unknown function	2 <i>E</i> -71	+	_f	+
bvORF23	At5g48620	Disease resistance protein (CC-NBS-LRR <sup>g</sup> class) family	e-107	+	+	+
bvORF24	At5g51740	Peptidase M48 family protein	8e-06	+	+	+
bvORF25	At5g35450	Disease resistance protein (CC-NBS-LRR class) family	<i>e</i> -100	+	+	+
bvORF26	At1g58390	Disease resistance protein (CC-NBS-LRR class) family	<i>e</i> -107	+	+	+
bvORF27	At2g04620	Cation efflux family protein	<i>e</i> -136	+	+	+
bvORF28	NA	No hit	NA	_	_	-
bvORF29	NA	No hit	NA	_	-	_
bvORF30	At5g23450	LCBK1, ATLCBK1, etc. (a sphingosine kinase)	0	+	+	+
bvORF31	At4g27870	Vacuolar iron transporter (VIT) family protein	2e-31	+	+	+
bvORF32	At4g27870	Vacuolar iron transporter (VIT) family protein	7e-35	+	+	+
bvORF33	At3g02580	Brassinosteroid biosynthetic enzyme		+	+	+
bvORF34	At5g24680	Peptidase C78, ubiquitin fold modifier-specific peptidase 1/2	2e-39	+	+	_
bvORF35	At3g49590	Autophagy-related protein 13	4e-96	+	+	+
bvORF36	At5g24660	RESPONSE TO LOW SULFUR 2 (LSU2)	2e-13	ND	ND	ND
bvORF37	NA	No hit	NA	ND	ND	ND
bvORF38	NA	No hit	NA	ND	ND	ND
bvORF39	At5g24650	Mitochondrial import inner membrane translocase subunit Tim17/Tim22/Tim23 family protein	5e-65	ND	ND	ND
bvORF40	At5g24630	BRASSINOSTEROID-INSENSITIVE4 (a protein that forms part of the topoisomerase VI complex)	3e-36	ND	ND	ND
bvORF41	At5g24620	Pathogenesis-related thaumatin superfamily protein	2e-76	ND	ND	ND

<sup>a</sup> Summary of Figure S1.

<sup>b</sup> Descriptions from TAIR (http://www.arabidopsis.org/).

<sup>c</sup> No data.

<sup>d</sup> Not applicable.

<sup>e</sup> Detected.

<sup>f</sup>Not detected.

<sup>g</sup> N-terminal coiled-coil domain (CC), central nucleotide-binding site domai (NBS) and C-terminal leucine-rich repeat (LRR).

RF and RFL by phylogenetic analysis (File S2) and found that bvORF16 clustered together with at5g42310 (labeled as At\_CRP1 in File S2) but not with any PPR-type RF or RFL proteins. Therefore, *bvORF16* is an atypical *Rf* candidate.

No genes in the 383,367-bp sequence exhibited homology to mitochondrial aldehyde dehydrogenase, glycine-rich protein, or retrograde regulated male sterility protein, which were encoded by maize *Rf2a*, rice *Rf2*, or rice *Rf17*, respectively (Table 1).

# The Oma1-Like gene was associated with NK–198-specific transcripts

We previously reported that a 7.0-kbp *Hin*dIII fragment that had been subcloned from 37O9 (a BAC clone overlapping with the 5A3, 9C23, and 33E19) detected specific transcripts



**Figure 2** Images of fluorescent signals obtained from transient expression tests. (A–F) Images of epidermal cells of Welsh onion sheath. (G and H) Images of epidermal cells of onion bulb scales. Bars, 50  $\mu$ m. A and D are green fluorescence images of bvORF16-GFP; B and H are red fluorescence images of mitochondria-targeted RFP; C is a merged image of A and B; E is a red fluorescence image of plastid-targeted RFP; F is a merged image of D and E; G is a green fluorescence image of bvORF20-GFP; and I is a merged image of G and H.

in flower buds of NK-198 but not of the CMS line, TK-81mm-MS (i.e., [S]rf1rf1) (Hagihara et al. 2005b). During our sequence analysis, we noticed that the 7.0-kbp HindIII fragment included the coding sequence of one of the quadruplicated genes, bvORF19, that resembled yeast Oma1, a peptidase M48 family protein involved in quality control of mitochondrial membrane proteins (Kaser et al. 2003) (Table 1 and see Files S3, S4, and S5). To see whether NK-198-specific transcripts were homologous to bvORF19, RNA gel blot analysis was conducted using the 3'-UTR sequence of bvORF19 as a probe. Because of high sequence homology among bvORF18-bvORF21, the design of specific hybridization probes for bvORF18, bvORF19, bvORF20, and bvORF21 was infeasible. Therefore, our probe simultaneously detected transcripts of the four genes in NK-198 samples. A strong signal appeared in the lane corresponding to NK-198 flower buds, but was hardly seen elsewhere (Figure 3). This result was consistent with our previous results using the 7.0-kbp HindIII fragment of the NK-198 genome (Hagihara et al. 2005b).

Both RNA gel blot analysis and RT–PCR analysis (see above) revealed that at least one copy of the quadruplicated genes (*bvORF18–bvORF21*) was expressed in anthers of NK–198, but it remained unclear whether all copies were expressed. Multiple sequence alignment of the *bvORF18*- to *bvORF21*- coding regions revealed that *bvORF18* and *bvORF21* were identical at the nucleotide sequence level, and thus could not be distinguished from each other (Figure S5). On the other hand, the sequences from nucleotide ~478 to ~497 provided unique sequence tags for *bvORF19* and *bvORF20*, due to a microsatellite-like polymorphism and nucleotide

substitutions (Figure S5). Based on this observation, we set up an assay including direct sequencing of RT-PCR products to detect the sequence tags of the expressed copies. Before we conducted the expression assay, the genomic DNA of NK-198 was subjected to PCR amplification, targeting a region encompassing the polymorphic sites (Figure 4) with primers D-Fw and D-RV to obtain a control template. The sequencing electrophoregram of the control template with the sequencing primer Gre is shown in Figure 4. At polymorphic site 1, a C residue occurs in *bvORF18* and bvORF21, whereas T and A are found in bvORF19 and bvORF20, respectively. We next PCR amplified cDNA of NK-198 young anthers (i.e., predehiscence) with the primers D-FW and D-RV. An electrophoregram of the RT-PCR products was obtained using the sequencing primer Gre. The highest peak at site 1 was A, followed by C and T. At polymorphic site 2, the peak of T, indicative of bvORF20, was higher than that of the control (Figure 4), although this may not reflect a significant quantitative difference. These data indicated that all copies of bvORF18-bvORF21 were expressed in anthers.

### bvORF20 restored partial pollen fertility to CMS sugar beet

If one of the quadruplicated ORFs is the *Rf1* gene, we might expect that the ORF in question could restore pollen fertility when transferred to Owen CMS plants. Sugar beet is known to be quite recalcitrant to regeneration following genetic transformation (Skaracis 2005). One of the present authors also found that regeneration in sugar beet was highly genotype



**Figure 3** RNA gel blot analysis of the 3'-UTR of *bvORF19* hybridized with total RNAs from flower buds (F), leaves (L), and roots (R) of NK–198, and from flower buds of TK–81mm–O. Sizes of signal bands are indicated in kilobases. Images in the bottom row show ethidium-bromide (EtBr-) rRNA after gel electrophoresis.

dependent, and a Japanese breeding line, NK-219mm-CMS, had a good shoot regeneration response (H. Tamagake, unpublished data). To examine whether NK-198 actually acted as a restorer of NK-219mm-CMS, we crossed NK-219mm-CMS with NK-198. The F<sub>1</sub> progeny (11 plants) were all classified as "partial fertile"; nearly all pollen grains appeared to be well developed morphologically but their cytoplasm was scarcely stained with Alexander's dye (Figure 5, A-C). Because this phenotype could be clearly distinguished from the completely sterile phenotype of NK-219mm-CMS plants (almost all microspores were aborted at an early stage of microsporogenesis and the exine was poorly developed), we concluded that NK-198 Rf1 restored partial fertility to the NK-219mm-CMS plants, although NK-198 Rf1 restored almost complete fertility to two other sugar beet lines, TK-81mm-CMS and TK-76mm-CMS (Hagihara et al. 2005a). Notably, the effect of NK-198 Rf1 is influenced by the nuclear genetic background (see the result using sugar beet line I-12CMS(R) in Hagihara et al. 2005a).

To test our hypothesis with transgenic plants, the genomic DNA fragment containing the protein-coding region and its 5' upstream (2 to 2.5 kbp in length) and 3' downstream regions (~500 bp) of *bvORF18*, *bvORF19*, *bvORF20*, or *bvORF21* were separately inserted into binary vectors. The resultant constructs were named pBVORF18, pBVORF19, pBVORF20, and pBVORF21, respectively. These constructs were subsequently introduced into NK–219mm–CMS calli by *Agrobacterium*-mediated transformation. The calli resistant to bialaphos herbicide, a phenotype conferred by the selectable marker on the T-DNA, were transferred to a regeneration medium. The regenerated sugar beet plants contained the bialaphos-resistance gene as shown by PCR analysis using primers BAR5 and BAR6 (data not shown).

We obtained 10 independent transgenic sugar beet plants transformed with pBVORF20, of which 8 exhibited partial



**Figure 4** Polymorphic sites in the PCR targets of the quadruplicated genes in NK–198 and electrophoregrams obtained by direct sequencing. The original electrophoregrams were converted to complementary images on the sequencing platform (ABI3130). Red, green, and black lines indicate the signal peaks of adenine, thymine, and cytosine, respectively. Numbers of nucleotides correspond to the sequence alignment shown in Figure S5.

fertility (Figure 5G). This partial-fertile phenotype was indistinguishable from that of the F<sub>1</sub> progenies of NK–219mm-CMS × NK–198 (Figure 5C). To ascertain the cosegregation of fertility restoration with the transgene, a transgenic plant carrying pBVORF20 was pollinated with the TA–33–O line, which had a maintainer genotype. The 14 F<sub>1</sub> plants were either male sterile (8 plants) or partial fertile (6 plants) (Figures S7 and S8). The bialaphos-resistance gene was found to cosegregate with the partial-fertility phenotype (Figure S8).

By contrast, three, nine, and eight transgenic plants were obtained carrying the pBVORF18, pBVORF19, and pBVORF21 constructs, respectively, and they all exhibited complete male sterility, not partial fertility (Figure 5, E, F, and H). These experiments strongly indicated that *Rf1* most likely corresponded to *bvORF20*.

#### Intracellular localization of bvORF20

The TargetP and Predotar programs predicted that bvORF20 proteins would be localized in mitochondria (scores: TargetP, 0.847; Predotar, 0.85). We constructed chimeric *GFP* genes with 55 N-terminal amino acid residues of bvORF20 at their 5' ends. The plasmid carrying the chimeric *GFP* genes was bombarded into epidermal cells of onion bulbs. The green fluorescent signals matched well with the red signals from the mitochondrial marker construct, pMt-R, which was cobombarded (Figure 2, G–I), confirming that *bvORF20* encodes a mitochondrial protein.

#### Organization of the rf1 allele

Alteration(s) in nucleotide sequence was expected in the *rf1* allele. Using a probe of the 3'-UTR sequence of *bvORF19*,



**Figure 5** Photographs of anther contents from transgenic and control sugar beets. A–H are images of Alexander's staining. Bars, 20  $\mu$ m. (A) Anther contents of a maintainer line, NK–219mm–O. (B) Anther contents of a CMS line, NK–219mm–CMS. (C) Anther contents of an F<sub>1</sub> plant (NK–219mm–CMS × NK–198). (D) Anther contents of a transgenic sugar beet transformed with the pMDC123 vector. (E–H) Anther contents of transgenic sugar beets transformed with pBVORF18–pBVORF21, respectively.

which is highly conserved among bvORF18-bvORF21, a lambda-phage genomic library of a maintainer line, TK-81mm-O (Matsuhira et al. 2007), was screened, and five recombinant phages were obtained. Restriction mapping of the five clones using EcoRI and XbaI enabled us to assemble these clones into a contig of  $\sim$ 30 kbp (Figure 6). Gene mapping of bvORF17, bvORF20, and bvORF22 on the physical map was achieved by DNA gel blot analysis, and recombinant phage no. 10 was identified as containing all mapped genes. The insert in recombinant phage no. 10 was subjected to shotgun sequencing. A continuous 16,037-bp region was obtained after assembling 55 independent plasmid clones and subsequent correction of any ambiguities by sequencing PCR fragments encompassing the regions in question. In the 16,037-bp region, we found three homologous genes to bvORF17, bvORF20, and bvORF22, but none of the BNR copies (Figure 6). The order and orientation of the three ORFs was preserved between TK-81mm-O and NK-198, but the bvORF20-like gene was single copy (here-



**Figure 6** Physical map of the chromosomal region containing *bvORF17*, *bvORF20L*, and *bvORF22* of TK–81mm–O. Sizes of restriction fragments are shown in kilobase pairs. Five recombinant phage clones are indicated. Gene organization deduced from the nucleotide sequence of clone no. 10 is indicated below with a bar. Colors and directions of the horizontal arrows have the same meanings as in Figure 1.

after named *bvORF20L*). To examine the copy number of *bvORF20L*, the conserved 3'-UTR sequence was hybridized to total cellular DNA of TK–81mm–O. The number and size of the signal band was congruent with the sequence data (7.9, 7.0, 5.9, and 1.9 kbp for NK–198 and 5.9 kbp for TK–81mm–O) (Figure 7). A comparison of the amino acid sequences predicted from *bvORF20L* and its counterparts in NK–198 is shown in Figure S9. *bvORF20L* is similar to the copies of the quadruplicated genes in NK–198 but not identical to any one of them (see Figure S5 for comparison of nucleotide sequences). Homologies at the amino acid sequence level were 83–85% (*vs. bvORF18, bvORF19, bvORF20*, and *bvORF21*). A detailed organizational comparison of this genomic region between TK–81mm–O and NK–198 will be presented elsewhere.

### Discussion

The nucleotide sequence of a 383-kbp chromosomal region containing the Rf1 locus of sugar beet was determined. Forty-one potential genes were found in this region. On this basis the gene density was calculated to be 9.4 kbp/gene, which appeared quite rich, given that the sugar beet's entire genome is 758 Mbp (Arumuganathan and Earle 1991). This gene density would suggest a total number of sugar beet genes of more than 80,000, an apparent overestimation compared to the total gene numbers of other dicots such as Arabidopsis (25,498), black cottonwood (Populus trichocarpa Torr. & A.Gray; 45,555), and grapevine (Vitis vinifera L.; 30,434) (The Arabidopsis Genome Initiative 2000; Tuskan et al. 2006; Jaillon et al. 2007). Recently, Dohm et al. (2012) reported that a maximum average distance of 30-40 kbp between genes in the sugar beet genome could be assumed according to their physical mapping study. On the other hand, three TEs identified in this study occupied



**Figure 7** DNA gel blot analysis of the 3'-UTR of *bvORF19* hybridized with total cellular DNA from NK–198 (lane 1) and TK–81mm–O (lane 2). *Hind*III restriction endonuclease was used. Size markers are shown on the right (in kilobase pairs).

a total of 6% of the sequenced region, which is much less than in other sugar beet chromosomal regions (up to 41.6%) (Schulte *et al.* 2006).

The 383-kbp region that was sequenced in this study contained neither typical PPR-type *Rf* gene nor genes related to *Rf* genes from other plants such as maize *Rf2a*, rice *Rf17*, or rice *Rf2* (Cui *et al.* 1996; Fujii and Toriyama 2009; Itabashi *et al.* 2011). This finding suggests that fertility restoration in sugar beet CMS involves a novel mechanism. This interpretation is consistent with the previous observation that mitochondrial gene expression in sugar beet is apparently unchanged after fertility restoration (Yamamoto *et al.* 2005).

On the other hand, we found that introduction of bvORF20 as a transgene restored partial fertility to NK-219mm-CMS. A comparable level of fertility restoration was observed in F<sub>1</sub> plants of NK–219mm–CMS × NK–198. Although three other ORFs homologous to bvORF20 were encoded in the Rf1 locus, none was capable of restoring male fertility. Therefore, despite their similarity in amino acid sequences, it is unlikely that these three ORFs play a major role in fertility restoration. Compared to bvORF20, the amino acid sequence homology in bvORF18, bvORF19, or bvORF21 is 88-99%. It is possible that one or more of the differences in amino acid sequences is involved in the inability to restore pollen fertility. Additionally, bvORF20L, a bvORF20-related gene found in *rf1rf1* sugar beet, encoded an uninterrupted ORF. Homology of the bvORF20L amino acid sequence to bvORF20 was 83%, and the amount of bvORF20L transcripts was greatly reduced compared to Rf1 sugar beet. Either or both of the structural or transcriptional alterations might render *bvORF20L* an *rf1* allele.

As far as we know, *bvORF20* homologs (Oma1 group in File S5) are conserved in eukaryotes as single copy genes. For example, the yeast homolog *Oma1* is involved in the quality control of mitochondrial membrane proteins with more or less similar activity as that of the matrix AAA pro-

tease (Kaser et al. 2003). In mammals, Oma1 functions as a membrane potential-dependent protease, one of whose substrates is OPA1, a GTPase involved in mitochondrial fusion (Ehses et al. 2009; Head et al. 2009). However, bvORF20 appears to lack protease activity because its  $Zn^{2+}$ -binding motif in the peptidase M48 domain is His-Gln-Val-Gly-His instead of the conserved His-Glu-x-x-His (Figure S9 and Files S3, S4, and S5). The Glu-to-Gln substitution in this motif was shown to abolish protease activity in yeast Oma1 (Kaser et al. 2003). According to our database search, ORFs homologous to yeast Oma1 preserve the His-Glu-x-x-His motif (File S4). These observations lead us to hypothesize that the function of *bvORF20* may not be a protease. On the other hand, if the possible molecular chaperone-like properties of yeast OMA1 (Kaser et al. 2003) are conserved in bvORF20, the bvORF20 protein might interact directly with preSATP6. This protein-protein complex might alter the higher order structure of preSATP6 to make it inactive. Molecular analysis of bvORF20 function is underway.

Concerning the evolution of plant *Rf*, the tandem gene cluster of bvORF18, bvORF19, bvORF20, and bvORF21 is reminiscent of the organization of the Rf loci of petunia, radish, and rice, whose translation products are PPR proteins (Bentolila et al. 2002; Brown et al. 2003; Desloire et al. 2003; Kazama and Toriyama 2003; Koizuka et al. 2003; Akagi et al. 2004; Komori et al. 2004). The evolutionary significance of such gene clusters may lie in the increased allelic diversity (Touzet and Budar 2004). We should point out an additional similarity that, in both PPR-type Rf loci and the sugar beet Rf1 locus, not all copies but one or several of these are capable of restoring fertility. Therefore, it is possible that a common mechanism has played an important role in the evolution of plant Rfs. We are currently investigating the organizational diversity of Rf1 in B. vulgaris plants to see how these genes have evolved.

#### Acknowledgments

We thank the DNA Sequencing Facility of the Research Faculty of Agriculture, Hokkaido University, for technical assistance. This work was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology, Japan; a grant from the Program for Promotion of Basic Research Activities for Innovative Biosciences, Japan; and the Program for Promotion of Basic and Applied Research for Innovations in Bio-Oriented Industry.

### **Literature Cited**

- Akagi, H., A. Nakamura, Y. Yokozeki-Misono, A. Inagaki, H. Takahashi et al., 2004 Positional cloning of the rice *Rf-1* gene, a restorer of BT-type cytoplasmic male sterility that encodes a mitochondriatargeting PPR protein. Theor. Appl. Genet. 108: 1449–1457.
- Alexander, M. P., 1969 Differential staining of aborted and nonaborted pollen. Stain Technol. 44: 117–122.

- Arimura, S., and N. Tsutsumi, 2002 A dynamin-like protein (ADL2b), rather than FtsZ, is involved in Arabidopsis mitochondrial division. Proc. Natl. Acad. Sci. USA 99: 5727– 5731.
- Arabidopsis Genome Initiative, 2000 Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796–815.
- Arumuganathan, K., and E. D. Earle, 1991 Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep. 9: 208– 218.
- Barkan, A., M. Walker, M. Nolasco, and D. Johnson, 1994 A nuclear mutation in maize blocks the processing and translation of several chloroplast messenger-RNAs and provides evidence for the differential translation of alternative messenger-RNA forms. EMBO J. 13: 3170–3181.
- Barr, C. M., and L. Fishman, 2010 The nuclear component of a cytonuclear hybrid incompatibility in *Mimulus* maps to a cluster of pentatricopeptide repeat genes. Genetics 184: 455–465.
- Bentolila, S., A. A. Alfonso, and M. R. Hanson, 2002 A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. Proc. Natl. Acad. Sci. USA 99: 10887– 10892.
- Bosemark, N. O., 2006 Genetics and breeding, pp. 50–88 in *Sugar Beet*, edited by A. P. Draycott. Blackwell Publishing, Oxford.
- Boutry, M., A. M. Faber, M. Charbonnier, and M. Briquet, 1984 Microanalysis of plant mitochondrial protein-synthesis products: detection of variant polypeptides associated with cytoplasmic male-sterility. Plant Mol. Biol. 3: 445–452.
- Brown, G. G., N. Formanova, H. Jin, R. Wargachuk, C. Dendy *et al.*, 2003 The radish *Rfo* restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. Plant J. 35: 262–272.
- Budar, F., R. Delourme, and G. Pelletier, 2004 Male sterility, pp. 43–64 in *Biotechnology in Agriculture and Forestry: Brassica*, edited by E. C. Pua and C. J. Douglas. Springer-Verlag, Berlin.
- Budar, F., P. Touzet, and G. Pelletier, 2006 Cytoplasmic male sterility, pp. 147–180 in *Flowering and Its Manipulation*, edited by C. Ainsworth. Blackwell Publishing, Oxford.
- Burge, C., and S. Karlin, 1997 Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. 268: 78–94.
- Chase, C. D., 2007 Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions. Trends Genet. 23: 81–90.
- Chenna, R., H. Sugawara, T. Koike, R. Lopez, T. J. Gibson *et al.*, 2003 Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. 31: 3497–3500.
- Chiu, W., Y. Niwa, W. Zeng, T. Hirano, H. Kobayashi *et al.*, 1996 Engineering GFP as a vital reporter in plants. Curr. Biol. 6: 325–330.
- Chomczynski, P., and N. Sacchi, 1987 Single-step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction. Anal. Biochem. 162: 156–159.
- Cui, X. Q., R. P. Wise, and P. S. Schnable, 1996 The rf2 nuclear restorer gene of male-sterile T-cytoplasm maize. Science 272: 1334–1336.
- Curtis, M. D., and U. Grossniklaus, 2003 A Gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiol. 133: 462–469.
- Desloire, S., H. Gherbi, W. Laloui, S. Marhadour, V. Clouet *et al.*, 2003 Identification of the fertility restoration locus, *Rfo*, in radish, as a member of the pentatricopeptide-repeat protein family. EMBO Rep. 4: 588–594.
- Dewey, R. E., C. S. Levings III, and D. H. Timothy, 1986 Novel recombination in the maize mitochondrial genome produce a unique transcriptional unit in the Texas male-sterile cytoplasm. Cell 44: 439–449.

- Dohm, J. C., C. Lange, D. Holtgrawe, T. R. Sorensen, D. Borchardt et al., 2012 Paleohexaploid ancestory for Caryophyllales inferred from extensive gene-based physical and genetic mapping of the sugar beet genome (*Beta vulgaris*). Plant J. 70: 528–540.
- Doyle, J. J., and J. L. Doyle, 1990 Isolation of plant DNA from fresh tissue. Focus 12: 13–15.
- Ducos, E., P. Touzet, P. Saumitou-Laprade, P. Vernet, and J. Cuguen, 2001 Nuclear effect on mitochondrial protein expression of the CMS Owen cytoplasm in sugar beet. Theor. Appl. Genet. 102: 1299–1304.
- Ehses, S., I. Raschke, G. Mancuso, A. Bernacchia, S. Geimer *et al.*, 2009 Regulation of *OPA1* processing and mitochondrial fusion by m-AAA protease isoenzymes and *OMA1*. J. Cell Biol. 187: 1023–1036.
- Emanuelsson, O., H. Nielsen, S. Brunak, and G. von Heijne, 2000 Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. 300: 1005– 1016.
- Finn, R. D., J. Mistry, B. Schuster-Bockler, S. Griffiths-Jones, V. Hollich *et al.*, 2006 Pfam: clans, web tools and services. Nucleic Acids Res. 34: D247–D251.
- Fujii, S., and K. Toriyama, 2008 Genome barriers between nuclei and mitochondria exemplified by cytoplasmic male sterility. Plant Cell Physiol. 49: 1484–1494.
- Fujii, S., and K. Toriyama, 2009 Suppressed expression of *RET-ROGRADE-REGULATED MALE STERILITY* restores pollen fertility in cytoplasmic male sterile rice plants. Proc. Natl. Acad. Sci. USA 106: 9513–9518.
- Fujii, S., C. S. Bond, and I. D. Small, 2011 Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution. Proc. Natl. Acad. Sci. USA 108: 1723–1728.
- Hagihara, E., N. Itchoda, Y. Habu, S. Iida, T. Mikami *et al.*, 2005a Molecular mapping of a fertility restorer gene for Owen cytoplasmic male sterility in sugar beet. Theor. Appl. Genet. 111: 250–255.
- Hagihara, E., H. Matsuhira, M. Ueda, T. Mikami, and T. Kubo, 2005b Sugar beet BAC library construction and assembly of a contig spanning *Rf1*, a restorer-of-fertility gene for Owen cytoplasmic male sterility. Mol. Genet. Genomics 274: 316–323.
- Hallden, C., C. Lind, and I. M. Moller, 1992 Variation in mitochondrial translation products in fertile and cytoplasmic malesterile sugar-beets. Theor. Appl. Genet. 85: 139–145.
- Hanson, M. R., and S. Bentolila, 2004 Interactions of mitochondrial and nuclear genes that affect male gametophyte development. Plant Cell 16: S154–S169.
- Head, B., L. Griparic, M. Amiri, S. Gandre-Babbe, and A. M. van der Bliek, 2009 Inducible proteolytic inactivation of *OPA1* mediated by the *OMA1* protease in mammalian cells. J. Cell Biol. 187: 959–966.
- Heitkam, T., and T. Schmidt, 2009 BNR a LINE family from Beta vulgaris - contains a RRM domain in open reading frame 1 and defines a L1 sub-clade present in diverse plant genomes. Plant J. 59: 872–882.
- Hjerdin-Panagopoulos, A., T. Kraft, I. M. Rading, S. Tuvesson, and N. O. Nilsson, 2002 Three QTL regions for restoration of Owen CMS in sugar beet. Crop Sci. 42: 540–544.
- Hu, J., K. Wang, W. Huang, G. Liu, J. Wang *et al.*, 2012 The rice pentatricopeptide repeat protein RF5 restores fertility in Hong-Lian cytoplasmic male-sterile lines via a complex with the glycinerich protein GRP162. Plant Cell 24: 109–122.
- Itabashi, E., N. Iwata, S. Fujii, T. Kazama, and K. Toriyama, 2011 The fertility restorer gene, *Rf2*, for Lead Rice-type cytoplasmic male sterility of rice encodes a mitochondrial glycinerich protein. Plant J. 65: 359–367.

- Jaillon, O., J. M. Aury, B. Noel, A. Policriti, C. Clepet *et al.*, 2007 The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449: 463–467.
- Jordan, D. R., E. S. Mace, R. G. Henzell, P. E. Klein, and R. R. Klein, 2010 Molecular mapping and candidate gene identification of the *Rf2* gene for pollen fertility restoration in sorghum [*Sorghum bicolor* (L.) Moench.]. Theor. Appl. Genet. 120: 1279–1287.
- Kaser, M., M. Kambacheld, B. Kisters-Woike, and T. Langer, 2003 *Oma1*, a novel membrane-bound metallopeptidase in mitochondria with activities overlapping with the m-AAA protease. J. Biol. Chem. 278: 46414–46423.
- Kazama, T., and K. Toriyama, 2003 A pentatricopeptide repeatcontaining gene that promotes the processing of aberrant *atp6* RNA of cytoplasmic male-sterile rice. FEBS Lett. 544: 99–102.
- Kitazaki, K., T. Kubo, H. Kagami, T. Matsumoto, A. Fujita *et al.*, 2011 A horizontally transferred tRNA<sup>Cys</sup> gene in the sugar beet mitochondrial genome: evidence that the gene is present in diverse angiosperms and its transcript is aminoacylated. Plant J. 68: 262–272.
- Klein, R. R., P. E. Klein, J. E. Mullet, P. Minx, W. L. Rooney *et al.*, 2005 Fertility restorer locus *Rf1* of sorghum (*Sorghum bicolor* L.) encodes a pentatricopeptide repeat protein not present in the colinear region of rice chromosome 12. Theor. Appl. Genet. 111: 994–1012.
- Koizuka, N., R. Imai, H. Fujimoto, T. Hayakawa, Y. Kimura *et al.*, 2003 Genetic characterization of a pentatricopeptide repeat protein gene, *orf687*, that restores fertility in the cytoplasmic male-sterile Kosena radish. Plant J. 34: 407–415.
- Komori, T., S. Ohta, N. Murai, Y. Takakura, Y. Kuraya *et al.*, 2004 Map-based cloning of a fertility restorer gene, *Rf-1*, in rice (*Oryza sativa* L.). Plant J. 37: 315–325.
- Kubo, T., and K. J. Newton, 2008 Angiosperm mitochondrial genomes and mutations. Mitochondrion 8: 5–14.
- Kubo, T., K. Kitazaki, M. Matsunaga, H. Kagami, and T. Mikami, 2011 Male sterility-inducing mitochondrial genomes: How do they differ? Crit. Rev. Plant Sci. 30: 378–400.
- Kurtz, S., J. V. Choudhuri, E. Ohlebusch, C. Schleiermacher, J. Stoye et al., 2001 REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 29: 4633–4642.
- Laser, K. D., and N. R. Lersten, 1972 Anatomy and cytology of microsporogenesis in cytoplasmic male sterile angiosperms. Bot. Rev. 38: 425–454.
- Lind, C., C. Hallden, and I. M. Moller, 1991 Protein synthesis in mitochondria purified from roots, leaves and flowers of sugar beet. Physiol. Plant. 83: 7–16.
- Matsuhira, H., H. Shinada, R. Yui-Kurino, N. Hamato, M. Umeda *et al.*, 2007 An anther-specific lipid transfer protein gene in sugar beet: its expression is strongly reduced in male-sterile plants with Owen cytoplasm. Physiol. Plant. 129: 407–414.
- Owen, F. V., 1945 Cytoplasmically inherited male-sterility in sugar beets. J. Agric. Res. 71: 423–440.
- Pelletier, G., and F. Budar, 2007 The molecular biology of cytoplasmically inherited male sterility and prospects for its engineering. Curr. Opin. Biotechnol. 18: 121–125.
- Pillen, K., G. Steinrücken, R. G. Hermann, and C. Jung, 1993 An extended linkage map of sugar beet (*Beta vulgaris* L.) including

nine putative lethal genes and the restorer gene *X*. Plant Breed. 111: 265–272.

- Sambrook, J., E. F. Fritsch, and T. Maniatis, 1989 Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Schmitz-Linneweber, C., R. Williams-Carrier, and A. Barkan, 2005 RNA immunoprecipitation and microarray analysis show a chloroplast pentatricopeptide repeat protein to be associated with the 5' region of mRNAs whose translation it activates. Plant Cell 17: 2791–2804.
- Schmitz-Linneweber, C., and I. Small, 2008 Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends Plant Sci. 13: 663–670.
- Schnable, P. S., and R. P. Wise, 1998 The molecular basis of cytoplasmic male sterility and fertility restoration. Trends Plant Sci. 3: 175–180.
- Schondelmaier, J., and C. Jung, 1997 Chromosomal assignment of the nine linkage groups of sugar beet (*Beta vulgaris* L.) using primary trisomics. Theor. Appl. Genet. 95: 590–596.
- Schulte, D., D. G. Cai, M. Kleine, L. J. Fan, S. Wang et al., 2006 A complete physical map of a wild beet (*Beta procumbens*) translocation in sugar beet. Mol. Genet. Genomics 275: 504–511.
- Skaracis, G. N., 2005 In vitro culture technique, pp. 247–255 in Genetics and Breeding of Sugar Beet, edited by E. Biancardi, L. G. Campbell, G. N. Skaracis, and M. de Biaggi. Science Publishers, Plymouth, UK.
- Small, I., N. Peeters, F. Legeai, and C. Lurin, 2004 Predotar: a tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics 4: 1581–1590.
- Staden, R., 1996 The Staden sequence analysis package. Mol. Biotechnol. 5: 233–241.
- Touzet, P., and F. Budar, 2004 Unveiling the molecular arms race between two conflicting genomes in cytoplasmic male sterility? Trends Plant Sci. 9: 568–570.
- Tuskan, G. A., S. Difazio, S. Jansson, J. Bohlmann, I. Grigoriev et al., 2006 The genome of black cottonwood, *Populus tricho*carpa (Torr. & Gray). Science 313: 1596–1604.
- Wang, Z., Y. Zou, X. Li, Q. Zhang, L. Chen *et al.*, 2006 Cytoplasmic male sterility of rice with boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. Plant Cell 18: 676–687.
- Williams-Carrier, R., T. Kroeger, and A. Barkan, 2008 Sequencespecific binding of a chloroplast pentatricopeptide repeat protein to its native group II intron ligand. RNA 14: 1930–1941.
- Xu, X. B., Z. X. Liu, D. F. Zhang, Y. Liu, W. B. Song *et al.*, 2009 Isolation and analysis of Rice *Rf1*-orthologus PPR genes co-segregating with *Rf3* in maize. Plant Mol. Biol. Rep. 27: 511– 517.
- Yamamoto, M. P., T. Kubo, and T. Mikami, 2005 The 5'-leader sequence of sugar beet mitochondrial *atp6* encodes a novel polypeptide that is characteristic of Owen cytoplasmic male sterility. Mol. Genet. Genomics 273: 342–349.

Communicating editor: S. Poethig

# GENETICS

Supporting Information http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.112.145409/-/DC1

### Unusual and Typical Features of a Novel Restorer-of-Fertility Gene of Sugar Beet (Beta vulgaris L.)

Hiroaki Matsuhira, Hiroyo Kagami, Masayuki Kurata, Kazuyoshi Kitazaki, Muneyuki Matsunaga, Yuko Hamaguchi, Eiki Hagihara, Minoru Ueda, Michiyo Harada, Aki Muramatsu, Rika Yui-Kurino, Kazunori Taguchi, Hideto Tamagake, Tetsuo Mikami, and Tomohiko Kubo

TABLE S1 Primers used in this s	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'-CCGTCGACATGAAATGGAGCTGCGTTG-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'-AAAAAGCAGGCTTTAGATCTGCCGTTGCACAACG-3'
	orf19-genomic 3' rv	5'-AGAAAGCTGGGTGTATCTGGGACCTGGATTGAG-3'
Amplification of bvORF20 for transgene construction	attB1-ORF20 prom	5'-AAAAAGCAGGCTGTAACAGAGGGTTCAAATTGCGG-3'
	orf20-genomic3'rv	5'-AGAAAGCTGGGTGGTCCTGGATTGAGGGTTAAC-3'
Amplification of bvORF21 for transgene construction	attB1-ORF21 prom	5'-AAAAAGCAGGCTGAACCTGAACTGAACTTATTGG-3'
	orf21-genomic3'rv	5'-AGAAAGCTGGGTTACCTGGGTCCTGGATTAAG-3'
Detection of bialaphos-resistance gene	BAR5	5'-CGAGACAAGCACGGTCAACTTC-3'
	BAR6	5'-AAACCCACGTCATGCCAGTTC-3'

	Condition o	famplification			NK198		Size of PCR
Name of ORFs	Anneal-	Extension	Nucleotide sequences of primers	Anthers	Leaves	Roots	products (bp)
	ing (°C)	EXTENSION		+ –	+ –	+ –	
byOPE12	60	1:00	5'-CTGATTTTGGACGGAGCTTGTTCG-3'				630
50010112	00	1.00	5'-TGCATTGTAGAAACACCCGCGTAG-3'			-	000
	62	1:30	5'-CCAGGGACAGGGAAGACCAAGAC-3'				1100
DVORFIS	02	1.50	5'-AGTCCTCCTTTCCACCCGACAC-3'				1100
	56	0.30	5'-ATCTCCACTTGAAGGGCCAG-3'				250
DVOR 14	50	0.50	5'-TTCTCGTCAGACGGACTGAG-3'		-	_	250
	56	0.20	5'-AGTTACCGTGGAGTTACTAGC-3'				300
DVORF15	50	0.30	5'-AGCACAGACTCGTTGCCACT-3'				300
	50	0.20	5'-AACATCTCCCTAGCCTTCCT-3'				970
DVORFIO	52	0.30	5'-CTGAATTCGTTTGCGTATAGT-3'				870
	56	1:00	5'-CAAGACTTGGTTCAATCAGCC-3'				550
DVORF17	50	1.00	5'-TTCTTTCTCGGCTTCAGCAGC-3'			- 200	550
bvORF18	56	0.30	5'-AAGGCATCCTCTCTTGCAAAA-3'				360
/19/20/21	50	0.50	5'-TGAATTGCACGTCCTGCTACA-3'				300
byOPE22	62	1.20	5'-GTGGCTCTCTCTAAACCGGCTTGT-3'				1400
	02	1.50	5'-CATGTTCAGCCCGACCCACGAA-3'				1400
h. ODE00	50	0.00	5'-CTATTGCGTGATCTTTGTGTTAGAA- 3'				440
DVORF23	53	0:30	5'-CTGGTATGTTATTATCAGAGTCAA-3'				410
	50	0.00	5'-TCGAATCTAACGCGGAGACA-3'				000
DVORF24	56	0:30	5'-TGCAGAGGGAGTCAAGTCAG-3'				230
	54	0.00	5'-ACAGGATTCGCTGGCCTTAA-3'				040
DVURF25	54	0:30	5'-TCAAAATTGGTCCTCACCAC-3'				240

	Condition of	of amplification				I	NK198			
Name of ORFs	Anneal-	Extension	Nucleotide sequences of primers	Ant	hers	Leav	/es	Roots	- c	Size of PCR products (bp)
	ing (°C)			+	-	+	-	+ –		
byORE26	56	1.30	5'-GATGGAAGGTACATGCACAC-3'							100
50010120		1.00	5'-CAATGCCACGCCAACTTTCC-3'					100		100
byOPE27	56	0.20	5'-AAGCGTCAGATCCTTAACCC-3'							260
	50	0.50	5'-ACTATTGAGGAACTCTGCTGC-3'							200
h:/ODE20	50	0.20	5'-CACCAATTTTAGGGGCTCTA-3'							120
DVURF28	52	0:30	5'-AAAAATCCAATCCAATAAGTCC-3'							120
byODE20	56	0.20	5'-TTCTCGAACCATATCCCACC-3'							150
DVORF29	00	0:30	5'-TGTGAAAGTCGAGAGCTAAGG-3'							150
byODE20	56	0.20	5'-ATATTAAACCCACGGTCCGG-3'							420
DVORF30	00	0.30	5'-ATGAGACAGTCGTCCCATAG-3'							430
	E A	1:00	5'-GGATCATACCTGAAGAGTGT-3'	-						570
DVORF31	54	1.00	5'-TAAGAAGACCATGCTCTTCC-3'					اللم		570
byOPE22	56	0.20	5'-TTGAACTTCCTAGACCTGGAGT-3'							500
DVORF32	00	0.30	5'-CACCGAGCTTCTTAAGTAGCATGT-3'			-				500
byOPE22	56	0.20	5'-ACACTTCTTAGGGTGACGAAG-3'							170
DVORF33	50	0.50	5'-TGTTGAAGCAGTGTGGGGTG-3'							170
	50	1:00	5'-TGGCAAAGGGGTTTTGACAC-3'							500
DVORF34	90	1:00	5'-GCAATTCCAGGATCAACATAGCAC-3'			-				000
	54	0.20	5'-TCTGATGTATCCACATCATCG-3'							220
DVORF35	54	0.30	5'-ATTAGATGCATCACGGTCTGG-3'	-				-		230

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).

TATGGGTAACTTATACATTGATAAGGGTAACTAATAGGGTAACTAATAGGATAATTTGAGCAACTAATGTT       100380         TTATATGTACATAGAACAAGTAATAACAACCCACGTACCGTGACTGAC	CAC	TTT	TGA	GCA	ACT	CAC	AAT	TTT	ATA	TAT	ACA	гта	CAA	GTA	ATT.	AAA	AA <u>T</u>	AAA	GTA	ГТА	100440
TTATATGTACATAGACAAGTAAATAAAATAACACACGTTACATTGACTTGACTTACACAAAGAAGA       100320         TAACTTATACATTGATAAAGGCAACTAATAAGACTAATAAGTTAAGCAACTAATGACAACTAAAGATTTTACAT       100260         GGCAACTAACGAATTTTATATAAAGACTAACTAATAGGTAACTAATAGGAACTAATAAGATAAAGATAAAGATAAAGATAAAGATAAAGATAAAGATAAAGATAAAGATAAAGATAAAGATAAAGATAAAGATAAAGAAATTA       100100         AGCAACTAACCAATTTATATAAAAAAAAAAAAAAAATATTAT	TAT	'GGG'	TAA	СТТ	ATA	CAT	TGA	TAA	GGG	TAA	СТА	ATA.	AGA	TAA	TTT	GAG	CAA	СТА	ATG	$\mathbf{T}\mathbf{T}\mathbf{T}$	100380
TAACTTATACATTGATAAAGACAACTAATAAGATAAGAT	TTA	TAT	GTA	CAT	AAC	AAG	TAA	ATA	ATA	ACA	CAC	GTT.	ACA	TTG	ACT	TGA	TTA	CAC	AAA	GGA	100320
TGACTTGGTTAGATAAAGGATAAACTTATAGATAGATAAAGGATAATAGAAAAGTAAAAGTAAAAGTAAAAGTAAAAAGTAAAAAGAAAGTAAAAAA	TAA	CTT	ATA	CAT	TGA'	TAA	AGA	CAA	CTA	ATA	AGA'	TAA	GTT	GAG	CAA	СТА	AAG	ATT	TTA	CAT	100260
AGCAACTAACAATTTTATATATATATATATATATATATAT	TGA	CTT	GGT	TAG	ATA	AAG	GAT	AAC	TTA:	TAG	ATA	GAT.	AAA	GGT	AAC	TAA	TAA	GAT	AAT	ГТG	100200
TATGGATAA       TAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	AGC	AAC	ГАА	CAA	TTT	TAT	ATA	TAC	ATT	ACA	AGT	AAT	ТАА	AAT	ACA	AAA	AGT	AAA	GTA	ГТА	100140
Argta Cattigattigattigattigattigattigattigatti	TAT	'GGA'	ΓΑΑ	TTA	AAA	AAA	AAA	AGT	AAA	ATA	TTA	TAT	GAG	GAA	СТА	TGC	AAA	TTC	CTA	СТА	100080
AGTAATCTATTTACTTTTAGATTAACTTÄAGCTTÄÄGGATTGGGATGGCCTATATGÄTGTTÄ 99900 ATÄÄGÄCATCÄCÄÄÄGTÄÄTTTTÄÄÄTTTÄÄÄTTTÄÄTTT	ATG	TAC	ATT	GAC'	TTG	GTT	CAG	TTA	AAC	TTG	ГТG	AAA	CAT	GTA	ATC	TTG	ТАА	TCC	TAG	IGG	100020
ATÀAGACATCACAACTAATCTCCCAACGTTGTTTATATTAATGTGCCATGTCATATGAAAAT 99900 GTCATAGACAAGTTCCAATAAGTCTAGATATGTCAAATTCAGTTACTAATTATATT 99840 GATTTATGTTCTTGTGTTGTTCCAGTAATCTACAATAGAATGGAAGAAGAAGAAGAAGAATTTGCTCC 99780 AACTCCTTGCATTGATGCACACTCCAACTCCACTCTGTGTTGATACAACTCCAACTAATGCTAC 99720 TCAAACTCCAAAGAATGGATCCAACTCCAACTCCACTCTGTGTTGATACAACTCCAACTAATGCTAC 99720 TCAAACTCCAAAGAATGGATCCCACTCCTGTGTTGATACAACTCCAACTCCACCACCACCACCACCACCACCACC	AGT	AAT	СТА	TTT	ACT	TTT	AGA	TTA	ACT	TAA	GCA'	TAG	TTT	GGG	ATG	GCC	TAT	ATG	ATG	гта	99960
GTCATAGACAAGTTCCAATAAGTCTAGATATGTCAAATTCAGTTTACGTTTACTAAATTATTATT 99840   GATTTATGTTCTTGTTGTGTGCCAGTAAGTCCAGTAACTCACAATAGAATGGAAGAGTGAAGAATTGCTCC 99780   AACTCCTTGCATTGATGCAACTCCAGTAACAACAAGATCCCACTCTGTGGATACAACTCCAACTAATGCTAC 99720   TCAAACTTCTAATGCCAACAACAAGATCCCACTCCTGTGTGATACAACTCCAACTACTGCACCACC 99600   R G S K E W I P C C P P E L K P T V G M      GCCTTTTGATTCTTTGTTGTGGTGGTACGTCTTGTGGTGGTATGGTCAAGTTGGCTCTTGTTGATCTTGTGGGGGGTATGGTCATGTTTGTGGGGGGTATGGCTACTGTGTGGGGGGTAAAGGTGTAAAGCTTGGTGTGGGGGGTAAGGTGGGGGGGG	ATA	AGA	CAT	CAC	AAC	TAA'	гст	CCA	ACG	TTG:	TTT	ATA	тта	ATG	TGC.	ATG	тса	TAT	GAA	AAT	99900
GATTTATGTTCTTGTTGTTCCAGTAACTACAATGGAAGATGGAAGAATTGCAAGAATTGCTC 99780   AACTCCTTGCATGTGTGTGCAACTCCAACTCCAACTCCTGTGTTGATACAACTCCAACTACTGCTAC 99720   TCAAACTTCTAATGCTCAAACAACAACAAGATCCCACTCCTGTGTGATACAACTCCAACTACTGCTACACC 99600   R G S K E W I P C C P P E L K P T V G M      GCCTTTTGATCTCTTGTTGATGGCTACTGGTTGTGATGGTTTTATAAGCTTATGCTCGGTTTTGATGGAAGAATTGGCTACTGGTACTGGTCACGTGTGTGAAGATGGCTACGTGAGAGATGGCTACGTGTGAAGATGGCTACGTGAGAGATGGCTACGTGTGAAGAGATGGCAGAGATGGCTACGTGTGAAGAGATGGCAGAGATGGCAGAGATGGCAAGAGATGGCAAGAGATGGCTACGTGTGAAGAGAGAG	GTC	ATA	GAC	AAG	TTC	CAA	TAA	GTC	TAG	ATA	TGT	CAA	ATT	тса	GTT	TAC	ТАА	ATT	ATT	ATT	99840
AACTCCTTGCATTGATTGATGCAACTCCAACTCCAACTCCTGTGTTGATACAACTCCAACTACTAATGCTAC       99720         TCAAACTTCTAATGCTCAACACTCCAACACACACAACACACTCCACTCGTCCAGGTCAACTCCAGGTCAACTCCACGTCAACACTCAACCACCACCACTGTTAAACCTCCAGGTCACGTCACGTCACCACCACTCTGTGGTAAACTCACGTCACTGTTTGATGGTTTTGATGGTATGGTTTTGATGGTATGGTTTTGATGGTAACGTTGGTCATGTTTGATGTTTGATGGTAACGATGGTCATGTTTGATGGAACGTCACGTGTGAGAAGGATGGTCATGTTTGATGTTTGATGGTAAACGATGGTCATGTTTACTTGAAGTAACGATGGTCATGTTTACTTGAAGTAACGAAGTAAACGATGGTCATGTTTACTTGAAGTAACGAAGTAAACGATGGTCATGTTTACTTGAAGTAACGAAGTAAACGATGGTCATGTTTACTTGAAGTAACGAAGTAAACGATGGTCATGTTTACTTGAAGTAACGAAGTAAACGATGGTCATGTTTACTTGAAGTAACGAAGTAAACGAAGTAAACGATGGTCATGTTTACTTGAAGTAAGGATGGTCATGTTTACTTGAAGTAACGAAGTAAACGAAGGAAG	GAT	'TTA'	TGT'	TCT	TGT	TGT	TCC	AGT	AAT	СТА	CAA	TAG.	ААТ	GGA	AGA	TGA	AGA	ATT	TGC'	гсс	99780
TCANACTTCTANTGCTCANCACGACGTCCACGTCCCTCGTCCACGTCACCACC       9960         R       G       S       K       E       W       I       P       C       C       P       P       E       L       K       P       T       V       G       M         P       F       D       S       L       V       D       G       I       P       F       V       K       P       T       V       G       M         CCCTTTTGATTCTCTTGTTGATGGTATTGAGGTATTGAGGTTTTAAAAGGATGGTCATGTTACTGGGGCTCATGTTACTGGAGAG       S       L       K       V       V       G       99600         P       F       D       S       L       V       D       G       I       E       F       Y       K       A       Y       A       R       F       C       G       M         F       V       E       R       L       A       T       E       K       K       D       G       H       V       Y       L       K       Y       P       P       P       P       P       P       P       P       P       P       P       P       P       P       P       P       P       P	AAC	TCC	TTG	CAT	TGA'	TGC	AAC	TCC	AAC	TCC'	TTG	TGT	TGA	TAC	AAC	тсс	AAC	TAA	TGC'	TAC	99720
R       G       S       K       E       W       I       P       C       C       P       P       E       L       K       P       T       V       G       M         GC       CTTTTGTTGTTGTTGTTGTGTGTTTGGTGTTTGGGTTTGGGTTTGGGTTTGGGTTGGCTGGTTTGGGTGGGTGGGGGG	тса	AAC	TTC	TAA'	TGC	TCA	AAC	AAC	AAG	ATC	CAC	ATT	CAC	TCC	TCG	тсс	ATG	CTA	CAC	ACC	99660
R G S K E W I P C C P P E L K P T V G M   GCCTTTTGTGTGTGTTGTTGTGGTGTTGGTGGGTTTGGGTTTGGGTGGGG	TAG	AGG	гтс	AAA	AGA	ATG	GAT	CCC	ГТG	CTG	CCC	FCC	TGA	GTT	AAA	ACC	TAC	TGT	GGG	TAT	99600
GCCTTTTGATTCTCTTGTTGATGGTATTGGTATTGATGGTTTTATAAAGCTTATGCTCGGTTTTGTGGGTTTGGTGGTGGTGGGGGGGG	R	G	s	K	Е	W	I	Ρ	С	С	Ρ	Ρ	Е	L	K	Ρ	т	v	G	М	
GCCTTTTGATTGATCTCTTGTTGATGGTATTGGTATTGATGGTTTTATAAGCTTATGCTGGTGTTTTGCGGTTTTGGGTGGTGGTGGTGGTGGTG																					
P F D S L V D G I E F Y K A Y A R F C G   TATTEGAAAGATGGCAAGATGGCAAGATGGCAAGAGAGAGAGAGA	GCC	TTT	IGA'	TTC	TCT	TGT	TGA'	TGG	TAT'	TGA	GTT	ГТА	ТАА	AGC	TTA	TGC	TCG	GTT	TTG:	IGG	99540
TTTTTGTGAAAGATTGCTACTGAGAAAAGAAAAGAAAAG	Ρ	F	D	S	L	v	D	G	I	Е	F	Y	К	А	Y	А	R	F	С	G	
F       V       E       R       L       A       T       E       K       C       A       T       E       K       C       A       A       T       E       K       C       A       A       T       E       K       C       A       A       T       E       K       C       A       A       T       E       K       C       A       T       E       K       C       A       T       E       K       C       A       T       E       K       C       A       C       A       Y       C       A       Y       C       A       K       Q       G       F       K       E       D       G       E       S       K       A       K       Y       P																					
F       V       E       R       L       A       T       E       K       K       D       K       D       G       H       V       Y       L       K       Y         TATTTATTGTAATAACAAGGATTTAAAGAAGATGGTGGAGAGAAGGTGAAGGTGAAGGTGTAAAGCAAGGAGTGAAGGTGTAAAGCAAGGATGTGAAGAGGTGTAAAGCAAGGAAGG	TTT	TGT	GGA	AAG	ATT	GGC	TAC'	TGA	GAA	AAA	AGA'	TAA	GGA	TGG	TCA	TGT	тта	CTT	GAA	GTA	99480
TATTTATTGTATATAACAAGGATTTAAAGAAGATGGTGAGAGTAAAGCAAGGTAAAGCAAGAGTAAAC       99420         I       Y       C       N       K       Q       G       F       K       E       D       G       E       S       K       A       K       S       K       P         TATAACATGCTCTAGTTCTCGTAGATGGAAGAGTGTAAATCGTGCTGGTGGTGGAGAGAGTGTAAACGTAGGAGAAATCCGTGGTGGAAAATCGTGGTGGAAAATCCGTGGTGGAAAATCCTGGTGGAAAATCCTTGGAAAATCCTTGAATCGCAAAGATCCGCAAGAATCCGCAAGAAATCCTTGGAAAATCCTTGAATCGCAAAATCGCAATAA       99300       G       L       R       K       R       S       V       N       R       A       G       C       Q       A       R       I         AGGTTTGAGAAAACGTAGTGATGGAAGAAATTCATGGTAGAAATTCTTTTTCATGAAAATCCTTAGAAATTCCTGAAAATTCCTGAAAATTCCTGAAAATTCCTGAAAATTCCTGAAAATTCCTGAAAATTCCTGAAAATTCCGAAAATTCCGAAAATTCTCGAAAATTCCGAAAATTCTGAAAATTCCGAAAATTCTGAAAATTCTGAAAATTCCGAAAATTCTGAAAATTCTGAAAATTCTGAAAATTCTGAAAATTCTGAAAATTCTGAAAATTCTGAAAATTCTGAAAATTCTGAAAATTCTGAAAATTCTGAAAATTCTGAAAATTCTGAAAATTCTGAAAATTCTGAAAATTCTGAAAATTCTGAAAATTCTGAAAAATTCTGAAAATTCTGAAAATTCTGAAAATTCTGAAAATTCTGAAAATTCTGAAAATTCTGAAAATTCTTGAAATTCTGAAAATTCTTGAAAATTCTGAAAATTCTTGAAAATTCTTGAAAATTCTTGAAAATTCTTGAAAATTCTTGAAAAATTCTTGAAATTCTTGAAAATTCTTGAAAATTCTTGAAAATTCTTGAAAATTCTTTGAAAATTCTTGAAAATTCTTGAAAATTCTTGAAAATTCTTGAAAATTCTTGAAAATTCTTGAAATTCTTGAAATTCTTGAAATTCTTGAAAATTCTTGAAATTCTTGAAATTCTTGAAATTCTTGAAATTCTTGAAATTCTTGAAATTCTTGAAATTCTTGAAATTCTTGAAATTCTTGAATTATATTCTAAGAATTCTTGAATTATATTCTTGAAATTCTTGAATTATATTCTTGAAATTCTTGAATTATATTCTTGAAATTCTTGAATTATATTCTTGAAATTCTTGAATTATTCTTGAATTATTCTTGAATTATTCTTGAAATTCTTGAAATTCTTGAAATTCTTGAATTATTATTGAAATTCTTGAATTTCTTGAATTTGAAATTCTTGAATTA	F	V	Е	R	L	А	т	Е	K	K	D	K	D	G	Н	v	Y	$\mathbf{L}$	K	Y	
I       Y       C       N       K       Q       G       F       K       E       D       G       E       S       K       A       K       S       K       P         TATTAACATGCTCTAGTTCTCGTAGAGAGTGTAAAGAGTGTAAACGTGTGTAAACGTGCTGGTTGTCAAGGAGGAT       P																					
I       Y       C       N       K       Q       G       F       K       E       D       G       E       S       K       A       K       S       K       P         TATAACATGCTCTAGTCCTAGTCCTAGTCCTAGTCGTGGTGTGTGT	TAT	'TTA'	L.L.	TAA'	TAA	ACA	AGG	ATT	TAA	AGA	AGA	rgg	TGA	GAG	TAA	AGC	AAA	GAG	TAA	ACC	99420
I       T       C       S       S       R       K       R       S       V       N       R       A       G       C       Q       A       R       I       99360         I       T       C       S       S       S       R       K       R       S       V       N       R       A       G       C       Q       A       R       I       99360         G       L       R       K       R       S       D       G       K       F       M       V       Y       L       F       H       E       S       H       N         CCATGTATTTGCCACTCCCAAAGCATGCATGCATGTCATGTGTAAAATTCCTGAAACTTGAACTTGAACTTGAACTTGAACTTGAACTTGAACTTGAACTTGAACTTGAACTTGAACTTGAACTTGAACTTGAACTTGAACAAAATCTTT       99240         H       V       F       A       T       P       K       S       M       H       F       L       K       N       S       R       N       L       T       L       99240         H       V       F       A       T       P       K       S       M       H       F       L       K       N       S       R       N       L       T       L <td>I</td> <td>Y</td> <td>С</td> <td>N</td> <td>K</td> <td>Q</td> <td>G</td> <td>F</td> <td>K</td> <td>Е</td> <td>D</td> <td>G</td> <td>Е</td> <td>S</td> <td>K</td> <td>А</td> <td>K</td> <td>S</td> <td>K</td> <td>Ρ</td> <td></td>	I	Y	С	N	K	Q	G	F	K	Е	D	G	Е	S	K	А	K	S	K	Ρ	
TATAACATGCTCTAGTTCTCGTAAAGAAGTGTAAATCGTGCTGCTGTGTCAAGCAAG																					
I T C S S S S R K R S V N R A G C Q A R I AGGTTTGAGAAAACGTAGTGGGGAAAATTCATGGTATATCTTTTCATGAATCGCATAA 99300 G L R K R S D G K F M V Y L F H E S H N CCATGTATTTGCCACCCCCAAAGCATGCATGCATTTTCTTGAAACTTCGAAACTTCGAAACTTCAAAATTC H V F A T P K S M H F L K N S R N L T L TGCTCACAAGAAGTTCATATTCATGTAATTCAAGATTGAATGTTGGACCAAAATCTT A H K K F I F D N S R L N V G P N K S F TAGATTGATAAAGGCCATGTAGGAGGATATGAGAATGTAGGGCCGTCATTGTTGATT 99120 P L L K F H V G G C Y F N V G A S L V D F	TAT	AAC	ATG	CTC	TAG'	TTC	TCG'	TAA	AAG	AAG'	TGT	AAA	TCG	TGC	TGG	TTG	тса	AGC	AAG	GAT	99360
AGGTTTGAGAAAACGTAGTGATGGAAAATTCATGGTATATCTTTTTCATGAATCGCATAA 99300 G L R K R S D G K F M V Y L F H E S H N CCATGTATTTGCCACTCCCAAAAGCATGCATTTTCTTAAAAATTCTCGAAACTTGACTCT 99240 H V F A T P K S M H F L K N S R N L T L TGCTCACAAGAAGTTCATATTTGATAATTCAAGATTGAATGTTGGACCAAACAAA	I	т	С	s	s	s	R	K	R	s	v	N	R	А	G	С	Q	А	R	I	
AGGTTTGAGAAAACGTAGTAGTAGGAGGATATGAGGGGGGGG																					
G L R K R S D G K F M V Y L F H E S H N CCATGTATTTGCCACTCCCAAAGCATGCATTTTCTTAAAAATTCTCGAAACTTGACTCT 99240 H V F A T P K S M H F L K N S R N L T L TGCTCACAAGAAGTTCATATTTGATAATTCAAGATTGAATGTTGGACCAAACAAA	AGG	TTT	GAG	AAA	ACG	TAG'	TGA'	TGG	AAA	ATT	CAT	GGT.	АТА	тст	TTT	тса	TGA	ATC	GCA	ГАА	99300
CCATGTATTTGCCACTCCCAAAAGCATGCATTTCTTAAAAATTCTCGAAACTTGACTCT 99240 H V F A T P K S M H F L K N S R N L T L TGCTCACAAGAAGTTCATATTTGATAATTCAAGATTGAATGTTGGACCAAACAAA	G	L	R	К	R	s	D	G	К	F	М	v	Y	L	F	н	Е	s	н	N	
CCATGTATTTGCCACTCCCAAAAGCATGCATTTCTTAAAAATTCTCGAAACTTGACTCT 99240 H V F A T P K S M H F L K N S R N L T L TGCTCACAAGAAGTTCATATTTGATAATTCAAGATTGAATGTTGGACCAAACAAA																					
H V F A T P K S M H F L K N S R N L T L TGCTCACAAGAAGTTCATATTTGATAATTCAAGATTGAATGTTGGACCAAACAAA	CCA	TGT	ATT	TGC	CAC	TCC	CAA	AAG	CAT	GCA'	TTT	гст	ТАА	AAA	TTC	TCG	AAA	CTT	GAC'	гст	99240
TGCTCACAAGAAGTTCATATTTGATAATTCAAGATTGAATGTTGGACCAAACAAA	Н	V	F	А	т	Ρ	K	S	М	Н	F	L	K	N	s	R	N	L	т	L	
A H K K F I F D N S R L N V G P N K S F TAGATTGATAAAAGAGCATGTAGGAGGATATGAGAATGTAGGGGGCGTCATTGGTTGATTT 99120 R L L K F H V G G V F N V G A S L V D F	TGC	TCA	CAA	GAA	GTT	CAT	ATT	TGA	TAA	гтс	AAG	ATT	GAA	TGT	TGG.	ACC	AAA	CAA	ATC	TTT	99180
TAGATTGATAAAAGAGCATGTAGGAGGAGGATATGAGAATGTAGGGGGGGG	A	Н	К	K	F	I	F	D	N	s	R	L	N	v	G	Р	N	К	s	F	
TAGATTGATAAAAGAGCATGTAGGAGGAGATATGAGAATGTAGGGGCGTCATTGGTTGATTT 99120																					
	TAG	ATT	GAT	AAA	AGA	GCA'	TGT	AGG	AGG	ATA	TGA	GAA	TGT	AGG	GGC	GTC	ATT	GGT	TGA'	TTT	99120
	R	L	I	к	Е	н	v	G	G	Y	Е	N	v	G	A	s	L	v	D	F	

5 SI

TA	AGA	AC	TTC	AG	FCG	AGA	<b>FGT</b>	TAA	AGC	TTA	CATA	ACAA	AGA	ГGТ	TGA	TGC	CGA	CAT	GTTI	ГGТ	99060
K	N		F	S	R	D	V	K	A	Y	I	Q	D	V	D	A	D	М	F	V	
AA	АТА	AT	TTC	'AA	AGA	AAA	GGC	AAC	TAG	<b>FAG</b>	rggi	rggi	AGG	GTT	TTT	CTT	IGA	CTA	FTG	ſGG	99000
N	N		F	К	Е	к	А	т	S	s	G	G	G	F	F	F	D	Y	С	G	
																				D	
AT	GAA	AA	TCG	GAC	ATT	TGA	СТА	GAG'	TTT	TTT	GGGG	CGGI	ATG	CCA	TTA	GTA	GGA	AAA	ACTI	\TT	98940
*																					
	E	N	R	Η	L	Т	R	V	F	W	A	D	A	I	S	R	K	N	Y	S	
СТ	СТТ	тт	TGG	TG	ATA	<b>FGG</b>	ГАТ	CAT	TTG	ATA	CAAC	CTT	ГТG	ATA	CCA	АТА	AAT	ATT	GTAT	ſGG	98880
	L	F	G	D	М	v	s	F	D	т	т	F	D	т	N	К	Y	С	М	v	
тт	СТТ	GC	ССС	AT	ГТА	CTG	GAG	ΓTG.	ATC	ATC	ATGO	GAA	AATO	GTG	TTA	CTT	ГТG	GTA	rggo	GCC	98820
	L	A	Ρ	F	т	G	v	D	Н	Н	G	K	С	v	Т	F	G	М	G	$\mathbf{L}$	
ΤА	СТТ	GC.	AAA	GGI	AAG	ATA	TAG	AAT	CTT	TCG:	TTTC	GGT	rgt:	ГТG.	AAT	GTT	TTT	TAA	AAGO	СТА	98760
	L	A	K	Ε	D	I	Ε	S	F	V	W	L	F	Ε	C	F	L	K	A	М	
		_																			
ΤG	GGT	AA	TTG	STC2	AAC	CTA	CTTO	GTC'	TCA'	TTA(	CTGA	ATC	AAG	ATG	CAG	CAA'	TGA.	AAC	AAG		98700
	G	N	С	Q	Р	т	C	Ц	T	т	D	Q	D	A	A	М	K	Q	А	T	
ጥጥ	CAA	<u>א</u> א	ъсл	ւաան	רכיתי	ኮጥል		יממי	ጥጥሮን	ልጥል	2201	ኮጥጥሪ	ברבי	ГСT	ccc	አጥልי	דריםי	TCA	<u>.</u>	AC	98640
11	GAA E	K K	V	. דד. ד	יבר. ד	K I I I	т	T	н	R	T.	с .	v	w W	н	Т	м	K I G	ĸ	V	90040
		IV.	v	1	1	ĸ	1	-	11	ĸ	Ц	C	v			-	11	ĸ	к	v	
ТG	CCG	GT.	AAA	AG	FAG	GTC	CAG	ATA	TGT	GTA	GAAG	CAAG	CGA	AGT	TTC	TTG	AGA	AAT	rga <i>i</i>	ATG	98580
	Р	v	K	v	G	Р	D	М	С	R	т	т	K	F	L	Е	к	L	N	A	
СТ	GTT	GТ	TTG	GG	ATA	GAG	ACC	ГТG	AGC	CAG	ATG	AAT	ГТG	ACA	AAG	GGT	GGA	ATT	CTGI	ſGA	98520
	v	v	W	D	R	D	L	Е	Ρ	D	Е	F	D	к	G	W	N	s	v	М	
ΤG	CGT	GA	ATI	TGC	GCT	TAG	AAG	ATG	ATG	GGT	GGTI	ГТА	CTG	ATA	TGT	TTA	ACA	TAA	GAC	ATA	98460
	R	Е	F	G	L	Ε	D	D	G	W	F	т	D	М	F	N	I	R	Η	М	
ΤG	TGG	AT	CCC	TTC	CTT	ACT	TTC	GAA	ATC	TTT	<b>FCA</b>	rgg(	GTG	GTA	TTT	TGA	GGT	CCA	CAC	AGA	98400
	W	Ι	Ρ	S	Y	F	R	Ν	L	F	М	G	G	I	L	R	S	т	Q	I	
mm	mar	<b>C</b> 7	<b>~</b>		N C 7 -	۰ <i>م</i> م			m <i>c</i>	<b>a</b> mm <sup>,</sup>	<b>1</b> 000	nm » /	<b>.</b>	<u>م سرم</u>	<b>~</b> ~ ~ ~	<u>م</u> سرم	<b>1</b> 111		<b>n x</b>		00240
.T.T.	rca g	GA F	GTC S	דע. ד	ND2- N	NUAA	нст". F	ידידי דידידי	тСАО	т.	רטיז קרטי	ר בי בי AC ייי		ATG A		T	т	т	TAG' W	ם די די	90340
	0		5	تد	TN	TN	Ľ	г	Ŧ	ш	Ľ	Τ.	TN	п	TA	ш	ш	ш	v	ند	

H. Matsuhira et al., FIGURE S2

6 SI

AGTT	'ATG	GTT	TCG	GAT	AGA	ATC	AGC	TAT	GGA	TGC	TCA	AAG	ACA	TGC	CCA	AAA	CAA	АСТ	CA	98280
$\mathbf{L}$	W	F	R	I	Е	S	A	М	D	A	Q	R	Н	A	Q	N	К	L	N	
АСТС	AGA	TTC	ТАА	GAA	TTC	САТ	GCC	тCG	тст	ТАТ	TAC	тсс	тст	TCC	ттт	AGA	GAA	GCA	ΤG	98220
S	D	S	K	N	S	М	Ρ	R	L	I	Т	Ρ	L	Ρ	L	Ε	К	Н	A	
CATC	TCT	TGT	TTA	CAC	ACA	CAA	TAT	GTT	СТА	CAA	ATT	тса	GAG	AGA	GTT	тса	AAA	TGC	AA	98160
S	L	v	Y	т	Н	N	М	F	Y	K	F	Q	R	Ε	F	Q	N	A	I	
TTTI	'TAA	TTG	TGG	GGT	тта	CAA	AGT	ACA	AAT	AGA	GGA	AGC	TGT	TGA	GGA	GTT	TGA	AGT	ΤG	98100
F	N	С	G	V	Y	K	v	Q	I	Ε	Ε	A	v	Е	Ε	F	Ε	v	A	
CAGA	TAA	TAC	AAG	GAA	GAA	AAC	АТА	тса	TGT	GAC	TTT	ТАТ	тсс	TGA	TTC	тса	TGA	TTG	тт	98040
D	N	т	R	К	К	т	Y	Н	v	т	F	I	Ρ	D	S	Н	D	С	F	
TTTG	CTC	TTG	ТАА	GAT	GTT	TGA	АТС	CAT	GGG.	ААТ	ATT	ATG	тсg	GCA	TGT	GCT	TTT	TGT	GA	97980
С	s	С	K	М	F	Е	s	М	G	I	L	С	R	Н	v	L	F	v	I	
TAAA	AGG	GAA	GTT	TTT	GAC	TGA	ААТ	TCC.	AGA	GCA	ACA	TAT	ATT	GCA	TCG	GTG	GAC	ТАА	AG	97920
К	G	К	F	L	т	Е	I	Ρ	Е	Q	Н	I	L	Н	R	W	т	K	D	
ATGC	TTC	AAA	AAA	GCC	САТ	TTT	CGA	СТТ	TTG	TGA	GGA	СТТ	TGA	TGG	ТАТ	AGA	ААТ	AAA	ТА	97860
A	S	K	K	Ρ	I	F	D	F	С	Ε	D	F	D	G	I	Ε	I	N	K	
AGAA	GAA	AAA	AGT	TGT	TGG	GGA	тст	TTG	GTC	GAA	ATT	СТТ	стс	ATG	TGT	AAG	ССТ	TGT	ТG	97800
K	K	K	V	V	G	D	L	W	S	K	F	F	S	С	V	S	L	V	Е	
AAAA	TAA	CAC	AGA	CCA	тст	TGA	GTT	ATT	ATT	GGA	AAG	GTT	ATC	TGC	ттт	TGA	GGA	GGA	AA	97740
N	N	т	D	Н	L	Ε	L	L	L	Е	R	L	S	A	F	Ε	Е	Е	М	
TGAA	ACC	TGG	AAA	AGA	AAA	TGT	TGA	GCA	ACA	АТС	ТАА	AGA	CAA	GCA	TAT	TGA	GTT	GTT	CG	97680
K	Ρ	G	K	Ε	N	V	Е	Q	Q	S	K	D	K	Н	I	Е	L	F	V	
TTGG	TTC	ТАА	ТАТ	AGT	АТС	AGG	TGG	TAT	АСТ	тсс	тсс	AAA	CAA	GTC	ттс	AAA	CAA	AGG	AA	97620
G	S	N	I	v	S	G	G	I	L	Ρ	Ρ	N	K	S	S	N	K	G	S	
GTGG	TAC	GGG	AAA	GAG	AAA	GAA	AAG	TGA	тса	AGA	GAT	AGC	CAT	TGA	AGC	AAG	CAA	CAA	AA	97560
G	т	G	к	R	К	К	S	D	0	Е	Ι	А	Ι	Е	А	S	N	К	K	

7 SI

AGGAGAGACTGTGTAGATCATGTGGTCAACTTTCAACTCATGATAGTCGTAATTGTCCTG 97500

ERLCRSCGQLSTHDSRNCPD

ATAAAAAGAAAAACCTAGAATGAGGTAAGTTTAGAATAATATATTATTATACATCACTA 97440 K K K N L E \*

TTTTGCAGTTACATGGAGGATTGTGATACATGGTGATGACAAGAATCATGATGATCAAAA 97320 GGGGAGCAAGCAGATTTTACATCACACTTTTTGGGGGAATTTTTGGTTTGGTTTTATAAAGT 97260 GTTACCTACAAAAAATAAAGTGTATGTGTTCAAAGGAGCTGGCAAACATTTAAAAAATTAT 97200 GGATTATTCTTTTGAAACGATTTCACTTGAAGGATTTTGATGGCTTAAGCACACTTTTGC 97140 AGTCTATCTCAAAAATTGTAGGTTTTGTGAAACTATATTCCATGGACACTGCAATTGATA 97080 TTGTAATTGGAAAGAACTTGATTATTCTGTAAAGTTTGCAAGTTATAACTTTTTCTTCTTC 97020 TCCACCTTTTTGTGTTACCTTTTAGTGCATTCTTTGACTAGGACACTGATTCATTGGTAA 96900 ΤΤΑΤΤΑΤGTAAAAACCCCCTATAAACAGTTACTAAAAAGAAGTTATAGTCAACCATACATT 96840 TACGAGAGAGTATTAATCTTTTAGTTTGATTTTATGTATTATTTTCAAATATCTTTTTA 96780 TATGTACATAACAAGTAAATAATTTCACACGTTACATTGACTTGATTACACAAAGGATAA 96720 CTTATACATTGATAAGGGTAACTAATAAGATAAGTTGAGCAACTAAAGATTTTTACATTG 96660 ACTTGGTTAGCCAAAGGATAACTTATAGATAGATAAGGGTAAATAATATGATAATTTGAG 96600 **CAACTAACAATTTTATATATACATTACAAGTAATTAAAATACAAGAAGTAAAGTATTATA** 96540 TGGATAACTTATAGATAGATAAGGGTAACTAATAAGATAATTTGAGCAACTAACAATTTT 96480 ΑΤΑΤΑΤΑΓΑΑΤΤΑCΑΑGTAATTAAAATACAAGAAGTAAAGTATTATATGGATAACTTATAC 96420 ΑΤΤGΑΤΑΑGGGTAACTAATAAGATAATTTGAGCAACTAACGATTTTATATGTACATAACA 96360 AGTAAATAATAACACACGTTACATTGACTATCTCCTTTCATCTTATCAAATAAAAAATA 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.

CTC	ATG	ACG	TCA	AGA	TCT	CAA	GTC	TAT	AAA	AAA	AAA	TTC	STTT	TAP	ATA	AG	ГСАА	CCT	TTG	11926
TGC	TGT	ACG	TCT	тст	TAT	TTT	'GAG	TCI	ACC	TTT	TGA	TT	<u>AT</u> TT	СТС	GATT	GAG	GTTA	AGC	TTG	11866
TAT	GTA	TGT	тст	CTT	СТА	TTG	AGT	TTT	'AAT	TTA	TTT	ATC	GACT	TGT	TTA	GT	AGGT	TAC	тта	11806
CTT	АСТ	TAT	GAT	ACG	TAT	TAC	AAT	GTC	CACI	CTC	GTC	TG	TTTG	AGA	AGA	AA	rgac	ATT	GTA	11746
ACA	CAC	ATC	АТА	AGT	AGT	CAT	TGC	ATT	TGT	'AAT	AGC	AA	CTT	AGG	GTAG	AGI	AGAG	AAT	GCC	11686
TAG	AGA	GAG	AGA	GAG	AGA	AAA	ACT	СТС	GAG	CGA	GCG	SAAG	GAGA	AGG	GAGA	ATC	GGAC	AAT	GGT	11626
																		М	V	
																		>0	RF-A	A
AAG	AAG	GAG	ACA	CCC	CCA	AGC	CAG	TAA	ACC	ACA	ACC	TAC	GAGC	СТТ	GAG	AAG	CAGC	CTT	CAT	11566
R	R	R	Н	Ρ	Q	А	S	K	Ρ	Q	Ρ	R	А	L	R	т	А	F	I	
							R	NA	rec	ogn	iti	on	mot	if	(pf	am	acc	ess	ion	PF00076)
AGA	TTT	ССТ	TCC	TCC	CAA	TAT	TGA	TAC	CCA	AAC	AAT	CCI	ACAA	CAI	TTAT	CAC	GTAG	ATA	TGG	11506
D	F	L	Р	Р	N	I	D	Т	Q	Т	I	Η	N	I	F	S	R	Y	G	
TGA	тст	GGA	GGA	CTT	AGT	GAT	ACC	AGC	CAAA	ACT	CCG	GAI	AAA	СТС	STGG	GC <i>I</i>	ACAA	ATA	CGC	11446
D	L	Е	D	L	V	I	Ρ	Α	K	$\mathbf{L}$	R	K	N	С	G	Η	K	Y	А	
ATT	CAT	ТАА	ATT	TTT	СТС	CAT	'GAA	TGC	TTT	ACT	CAA	TGO	CGAT	TAP	AGCA	GGI	AGAA	TGG	AAG	11386
F	Ι	K	F	F	S	М	N	А	$\mathbf{L}$	$\mathbf{L}$	N	Α	Ι	Κ	Q	Ε	N	G	R	
AAG	AAT	GGG	AAA	TTT	TTT	GAT	GCG	AGI	TAA	ccc	TGC	'AA	АТА	TGA	ACAA	ACI	AAGA	'CCC'	гсс	11326
AAG R	AAT M	GGG G	AAA N	TTT F	TTT L	'GAT M	'GCG R	AGI V	TAA N	CCC P	TGC A	AA K	ата У	TGA D	ACAA K	AC <i>I</i> Q	AAGA D	CCC, P	TCC P	11326
AAG R	AAT M	GGG G	AAA N	TTT F	TTT L	'GAT M	'GCG R	AGI V	TAA N	P	TGC A	'AA K	ата У	TGA D	ACAA K	AC <i>i</i> Q	AAGA D	CCC' P	TCC P	11326
AAG R CCA	AAT M .TAA	GGG G AAA	AAA N .CCA	TTT F .CTT	TTT L TCC	GAT M	GCG R	AGT V TAA	N N ACC	CCC P	TGC A	CAC	ата У Gaca	TGA D	ACAA K CTCA	ACA Q AAA	AAGA D AAAA	CCC P	ICC P GGT	11326 11266
AAG R CCA H	AAT <u>M</u> .TAA K	GGG G AAA N	AAA N CCA H	TTT F .CTT F	TTT L TCC P	GAT M AAA N	GCG R TCC P	AGI V TAA K	TAA N ACC P	CCC P P AAA N	TGC A TCA H	CAC R	AATA Y GACA Q	TGA D .GCC P	ACAA K CTCA Q	ACA Q AAA K	AAGA D AAAAA N	CCC P CCCC P	ICC P GGT V	11326 11266
AAG R CCA H	AAT M TAA K	GGG G AAA N	AAA N CCA H	TTT F .CTT F	TTT L TCC P	GAT M AAA N	GCG R TCC P	AGT V TAA K	TAA N ACC P	CCC P P AAA N	TGC A TCA H	CACAC R	AATA Y GACA Q	TGA D .GCC P	ACAA K CTCA Q	ACA Q AAA K	AAGA D AAAAA N	CCC P CCC P	TCC P GGT V	11326 11266
AAG R CCA H ACA	AAT M TAA K	GGG G AAA N TCA	AAA N CCA H	TTT F CTT F	TTT L TCC P TTG	GAT M AAAA N	GCG R TCC P AGA	AGT V TAA K	N N ACC P	CCC P AAAA N	TGC A TCA H	CAC R	AATA Y GACA Q AGGA	TGA D .GCC P .TG1	ACAA K CTCA Q CTCC	Q Q AAA K K	AAGA D AAAAA N	CCC P CCCC P P	ICC P GGT V CCA	11326 11266 11206
AAG R CCA H ACA Q	AAT M TAA K ATA Y	GGGG G AAA N TCA H	AAA N CCA H TCC P	TTT F CTT F AGC A	TTT L TCC P TTG W	GAT M AAAA N GAG R	GCG R TCC P AGA D	AGI V TAA K CCA	TTAA N ACCC R	CCC P CAAA N SATC S	TGC A TCA H SGTA Y	CAC R R TAZ	AATA Y GACA Q AGGA D	TGA D .GCC P .TG1 V	ACAA K CTCA Q CTCC S	Q Q AAAA K SGAA N	AAGA D AAAAA N ACCC P	CCC P CCC P AAAA N	ICC P GGT V CCA Q	11326 11266 11206
AAG R CCA H ACA Q	AAT M TAA K ATA Y	GGGG G AAA N TCA H	AAA N CCA H TCC P	TTT F .CTT F AGC A	TTT L TCC P TTG W	GAT M AAAA N GAG R	GCG R TCC P AGA D	AGI V TAA K ACCA H	N N NACC P NCCG R	CCC P AAAA N ATC S	TGC A TCA H GTA Y	K K CAC R R TA K	AATA Y GACA Q AGGA D	TGA D GCC P TG1 V	K K CTCA Q SCTC S	Q Q AAAA K CGAA N	AAGA D AAAA N ACCCC P	CCC P CCC P P SAAA N	TCC P GGT V CCA Q	11326 11266 11206
AAG R CCA H ACA Q AAT	AAT M TAA K ATA Y	GGG G AAA N TCA H AAT	AAA N CCA H TCC P	TTT F CTT F AGC A	TTT L TCC P TTG W	GAT M AAA N GGAG R TGT	GCG R TCC P AGA D	AGI V TAA K ACCA H	N ACC P ACCG R	CCCC P AAAA N ATC S CAA	TCA A TCA H SGTA Y TCC	CAA K ACAC R R ATA K	AATA Y GACA Q AGGA D	TGA D GCCC P TGI V V	ACAA K TCA Q TCTC S	AC <i>I</i> Q AAA <i>I</i> K SGA <i>I</i> N	AAGA D AAAA N ACCCC P	CCC P CCC P CCC N	ICC P GGT V CCA Q GAA	11326 11266 11206 11146
AAG R CCA H ACA Q AAT I	AAT M TAA K ATA Y ACC P	GGG G AAA N TCA H AAT I	AAA N CCCA H TCCC P CCCA H	TTT F CTT F AGC A CAC T	TTT L TCC P TTG W TGA D	GAT M AAAA N GGAG R TGT V	GCG R TCC P GAGA D TCCC P	AGI V TAA K ACCA H SACC P	N N AACC P ACCCG R R CAATI I	CCCC P ZAAA N SATC S CAAA N	TGC A TCA H Y TCC P	K CAC R R TAP K CTC S	AATA Y GACA Q AGGA D CAACC T	TGA D GCCC P TGI V CCAA K	ACAAA K CTCA Q CCTC S AACCC P	LACA Q LAAA K K SGAA N STAA N	AAGA D AAAA N ACCCC P ATAC T	CCC P CCCC P CCCG R	ICC P GGT V CCA Q GAA K	11326 11266 11206 11146
AAG R CCA H ACA Q AAT I	AAT M TAA K ATA Y ACC P	GGG G AAAA N TCA H AAT I	AAA N CCA H TCC P CCA H	TTT F CTT AGC A CAC T	TTT L TCC P TTG W TGA D	GAT M AAAA N GAG R TGT V	GCG R TCC P AGA D TCC P	AGT V TAA K CCA H P	N N AACC P ACCCG R R ZAAT I	CCCC P CAAA N SATC S CAA N	TGC A TCA H SGTA Y TCC P	K K ACAC R R K K CTC S	Y GACA Q AGGA D CAAC T	TGA D GCCC P TGI V CAA K	ACAA K CTCA Q CCTC S S AACC P	Q Q AAA K SGAA N TAA N	AAGA D AAAAA N ACCCC P ATAC T	CCCC P CCCCC P CAAAA N CCCGC R	ICC P GGT V CCA Q GAA K	11326 11266 11206 11146
AAG R CCA H ACA Q AAT I ACC	AAT M TAA K ATA Y ACC P ACC	GGG G AAAA N TCA H AAT I TCA	AAA N CCA H TCC P CCA H TCA	TTT F AGC A CAC T AAC	TTT L TCC P TTG W TGA D GAA	GAT M AAA N GAG R TGT V TTT	GCG R TCC P GAGA D TCCC P	AGT V TAA K CCA H ZACC P	N N AACCO P ACCCG R CAAT I I	CCCC P CAAA N CATC S CCAA N N CACC	TGC A TCA H SGTA Y TCC P	K K ACAO R K K CTO S	Y Y GACA Q AGGA D CAACC T	TGA D GCC P TGI V CAA K K	ACAA K Q CTCA Q CCTC S S AACCC P	Q Q AAAA K CGAA N TTAA N	AAGA D AAAAA N ACCCC P ATAC T T	CCCC P CCCCC P CAAAA N CCCGG R	TCC P GGT V CCA Q GAA K K	11326 11266 11206 11146 11086
AAG R CCA H ACA Q AAT I ACC P	AAT M TAA K ATA Y ACC P ACC	GGG G N TCA H AAT I TCA H	AAA N CCA H TCC P CCA H TCA Q	TTT F CTT F AGC A CAC T AAC T	TTT L TCC P TTG W TGA D GAAA N	GAT M AAA N GAG R TGT V TTTT L	GCG R TCC P GAGA D TCC P TCC P S	AGI V TAA K ACCA H ACCO P SCTC S	N N AACC P ACCG R CAAT I S	CCCC P CAAA N GATC S CAAA N CCAA N CCAA P	TGC A TCA H SGTA Y TCC P TAT I	K K CAC R K K CTC S CTC S CTC S CTC S CTC S CTC S C C C C	Y Y GACA Q AGGA D CAAC T T AATC S	TGA D GCCC P TGI V CCAA K K ZAAI I	ACAA K CTCA Q CCTC S CCTC S AACC P CCAT I	ACZ Q LAAZ K SGAZ N STAZ N N CCCC P	AAGA D AAAAA N ACCCC P ATAC T T CTAAA N	CCCC P CCCCC P CAAAA N CCCG R CCCA Q	ICC P GGT V CCA Q GAA K AAT I	11326 11266 11206 11146 11086
AAG R CCA H ACA Q AAT I ACC P	AAT M TAA K ATA Y ACC P ACC	GGG G AAAA N TCA H TCA H	AAA N CCA H TCC P CCA H TCA Q	TTT F CTT F AGC A CAC T AAC T	TTT L TCC P TTG W TGA D GAAA N	GAT M AAA N GAG R TGT V TTT L	GCG R TCC P AGA D TCC P TCC P S	AGT V TAA K ACCA H ZACC P SCTC S	N N ACCC R CAAT I S TTC S	ACCC P AAAA N ATC S CAAA N S ACCAA P	TGC A TCA H SGTA Y TCC P TAT I	K K CAC R K CAC K S CAC S CAC E	AATA Y GACA Q AGGA D CAAC T T AATC S	TGA D GCC P TGI V CAA K K AAI I	ACAA K CTCA Q CCTC S AACC P CCAT I	ACZ Q LAAZ K CGAZ N TAZ N N CCCC P	AAGA D AAAAA N ACCCC P ATAC T T CTAAA N	P ACCCC P ZAAAA N CCCG R R ACCCA. Q	ICC P GGT V CCA Q GAA K K AAT I	11326 11266 11206 11146 11086
AAG R CCA H ACA Q AAT I ACC P CCT	AAT M TAA K ATA Y ACC P ACC P TGA	GGG G N TCA H AAT I TCA H ACC	AAA N CCA H TCC P CCA H TCA Q TCT	TTT F CTT F AGC A CAC T T CAG	TTT L TCC P TTG W TGA D GAAA N TAC	GAT M AAA N GAG R TGT V TTTT L	GCG R TCC P GAGA D TCC P TCC P S CAT	AGI V TAA K ACCA H ACCA P SCTC S	N AACC P ACCG R CAAT I S CTTC S	ACCC P ZAAA N GATC S CAAA N CAAA N CAAA P AGA	TGC A TCA H SGTA Y TCC P TAT I I AAT	K ACAC R ATAZ K SCTC S CAGZ E	Y Y GACA Q AGGA D CAAC T XATC S	TGA D GCCC P TGI V CCAA K K CAAI I I GCA	ACAA K CTCA Q CCTC S AACCC P CCAT I ACCCG	ACZ Q AAAZ K SGAZ N STAZ STAC	AAGA D AAAAA N ACCCC P ATAC T T CTAAA N GGAT	P ACCCC P AAAAA N ACCCG R ACCCA. Q	ICC P GGT V CCA Q GAA K AAT I ITTC	11326 11266 11206 11146 11086 11026

TAG	GGT	ССТ	TGG	GGA	AGA	CAC	GGA	GAG	AAT	AAG	GGA	CCA	AGT	GGA	ACT	TGT	GGA	АСТ	AGA	10966
R	V	L	G	Е	D	Т	Е	R	I	R	D	Q	V	Ε	L	V	Е	L	Е	
GGG	CGA	TCA	GAT	TCT	TGC	CAT	CTC	AGG	GGA	GAA	AAA	TGA	AGA	GAT	ССТ	GGA	GTT	АСТ	GGA	10906
G	D	Q	I	L	A	I	S	G	Е	K	N	Е	Е	I	L	Ε	L	L	Е	
AAG	AAG	CGT	TAT	AGC	AGT	CGC	AAA	СТС	TTC	ATC	TCC	ATC	CAA	GAT	TAT	CCA	TGA	GCA	ТАТ	10846
R	S	V	I	A	V	A	N	S	S	S	Ρ	S	K	I	I	Н	Е	Η	I	
CTT	GGC	GGA	AGG	GGI	TAA	СТА	тст	GAA	GAT	TAA	ACC	сст	TGG	GGG	AAT	GCT	TCA	тст	TAT	10786
L	A	Е	G	V	N	Y	L	K	I	K	Ρ	L	G	G	М	L	Η	L	I	
CCA	GTT	CAA	TTC	GGI	TGA	AGA	AAA	GGA	TGA	CAT	GAT	AAA	AAG	CAA	ATG	GCT	TGA	ACG	ATG	10726
Q	F	N	S	V	Е	Е	K	D	D	М	I	K	S	K	W	L	E	R	W	
GTT	ССТ	GGA	GCT	'AAG	GGA	TGT	'GAA	ТАА	CGC	TAG	CAC	GGC	ATT	'ATG	GAG	GGA	GAT	GTG	GAT	10666
F	L	Е	L	R	D	V	N	N	A	S	Т	A	L	W	R	Е	М	W	I	
CAC	ААТ	ጥጥል	тGG	AGT	ידיככ	'Aጥጥ	GАT	CGC	ATG	GAG	ጥጥል	тGA		ጥጥጥ	тса	GAA	ልልጥ	тGG	ጥጥG	10606
Т	I	Y	G	V	P	L	I	A	W	S	Y	E	N	F	Q	K	I	G	C	10000
ጥልጥ	ል ጥጥ	CCC	GAG	a Cu	יכריד	י አ ጥ ር י	сст		ልጥል	ጥጥር	тсc	Слл	יככא	ጥጥል	ccc	CAC	ልርጥ	ጥሮ ል	ርጥጥ	10546
I	F	G	R	V	L	S	V	E	Y	S	R	M	D	Y	A	R	V	Q	L	10540
<u>አ</u> አጥ	CAC	лсл	ጥጥር	יייריי	יריתי	ירים	аст	ממיזי	መ አ አ	ccc	Сат	аст	ւաա	מיתיתי	ርርሞ	CCA	лсл	መል እ	ACC	10486
I	T	D	C	L	F	K	V	N	N	P	I	V	F	Y	V	E	D	K	P	10400
GТТ	таа	GAT	יתיתי	ידיGיד	ידאכ	AGA	AGA	СТТ	ידיGG	тст	тGG	тсс		тса	тGA	тсс	тсс	тGC	AAG	10426
F	K	I	F	V	T	E	D	F	G	L	G	P	N	Н	D	P	P	A	S	10120
ТАА	AGG	TAT	GCC	'AAA	TCC	сст	יכיד		TAG	АТТ	AGA	TTC	TGA	таа	СТС	GAA	TTC	GGA	АТС	10366
K	G	м	P	N	.100 P	L	F	Н	R	L	D	s	D	N	S	N	S	E	S	10000
	-		_		_	_	_			_	_	_						_		
СТС	TGA	TAA	AGA	TCC	ATT	GGA	TGA	TGA	TGA	TCG	TGA	TAG	TGA	CGA	CTG	GGA	TCC	тсс	GGG	10306
S	D	K	D	Ρ	L	D	D	D	D	R	D	S	D	D	W	D	Ρ	Ρ	G	
AGG	GGA	AAG	GTC	ACC	CCA	AAA	ACC	CCT		AAA	СТС	CGA	GTT	CAA	TGC	ATC	GGG	ΑΑΑ	TAC	10246
G	Е	R	s	P	Q	K	P	L	P	N	s	E	F	N	A	s	G	N	Т	•

TCC	AGG	AAT	СТС	GGA	CAT	TGG	AGT	ТАА	TAA	TGA	GAC	AAT	CAT	TAC	TTC	GCC	TAC	CAA	AGC	10186
Ρ	G	I	S	D	I	G	V	N	N	Ε	т	I	I	т	S	Ρ	Т	K	A	
TAG	TGC	CAA	GGT	стс	тсс	CAA	TGC	ААА	ACA	ААА	ACC	TCC	САТ	АТА	ccc	ТАА	ACC	TCC	CAA	10126
S	A	K	v	S	Ρ	N	A	K	Q	K	Ρ	Ρ	I	Y	Ρ	K	Ρ	Ρ	K	
AAC	TCA	ACT	GAA	СТТ	ТАА	TAC	ссс	ACC	ACG	TTC	ссс	AAG	тст	GCT	TTG	CAT	TGG	GAA	CTT	10066
Т	Q	L	N	F	N	Т	Ρ	Ρ	R	S	Ρ	S	L	L	С	I	G	N	L	
AAA	TCA	ACA	AAA	GTC	стс	стс	CCA	ACC	АСТ	TGA	АСТ	CCA	AAA	AGC	ccc	ACC	TTC	ACC	ATC	10006
N	Q	Q	K	S	S	S	Q	Ρ	L	E	L	Q	K	A	Ρ	Ρ	S	Ρ	S	
GAA	AAC	CTT.	ACC	СТТ	ccc	тсс	AAC	AAC	GAA	АСТ	GGG	СТС	ACC	TTT	TAG	ccc	TGA	тсс	AAC	9946
K	Т	L	Ρ	F	Ρ	Ρ	Т	Т	K	L	G	S	Ρ	F	S	Ρ	D	Ρ	т	
СТТ	TAA	ATA	ТАА	ТАА	TCC	ccc	САТ	стс	CCA	AAA	TAA	TAT	AAT	CAG	ccc	AAT	AAG	CCC	ATT	9886
F	K	Y	N	N	Ρ	Ρ	I	S	Q	N	N	I	I	S	Ρ	I	S	Ρ	L	
GGI	ccc	CAA	ACC	TGC	CCA	AAA	TAC	ACA	AAA	СТС	ссс	TAG	TTC	TAC	AAG	TCG	AAA	СТС	TCC	9826
v	Ρ	K	Ρ	A	Q	N	Т	Q	N	S	Ρ	S	S	Т	S	R	N	S	Ρ	
TTT	AAA	GCC	CAG	ССТ	CAA	TGA	CCA	AAG	СТТ	тсс	TTA	СТА	CAA	TCC	тст	GAT	CCA	CAC	TGA	9766
L	K	Ρ	S	L	N	D	Q	S	F	Ρ	Y	Y	N	Ρ	L	I	Η	Т	D	
TAA	TTC	CTT	TGG	ccc	GCT	АСТ	AAG	GAA	AGC	CCA	ATC	AAA	ATC	CCA	AAC	ТАА	GAC	ACT	CTC	9706
N	S	F	G	Ρ	L	L	R	K	A	Q	S	K	S	Q	Т	K	Т	L	S	
ATC	СТС	TCC	TTC	GAC	GTC	CAG	ссс	TTC	TAT	ccc	ссс	CGG	TTT	TGA	AGA	СТТ	сст	TCC	TCC	9646
S	S	Ρ	S	Т	S	S	Ρ	S	I	Ρ	Ρ	G	F	Ε	D	F	L	Ρ	Ρ	
ccc	тст	GAA.	AGC	CCA	TCA	TGA	AAA	AAG	GAG	ATT	ACA	AAA	ACG	ACT	GAA	GAA	AAA	TAA	AGC	9586
Ρ	L	K	A	Η	Н	Ε	K	R	R	L	Q	K	R	L	K	K	N	K	A	
CAA	AAA	CCG	ССТ	стс	стс	стс	стс	стс	CAA	TCC	ссс	ACC	тст	TCC	тсс	СТС	тсс	СТС	CCC	9526
K	N	R	L	S	S	S	S	S	N	Ρ	Ρ	Ρ	L	Ρ	Ρ	S	Ρ	S	Ρ	
AAA	CCC	GAA.	AAC	ATC	тса	TGA	GAA	CAC	TGC	СТС	GGA	ААТ	TAT	TGA	ATT	AGG	СТТ	GCA	АСТ	9466
N	Р	K	т	S	Н	Е	N	т	А	s	Е	I	I	Е	L	G	$\mathbf{L}$	Q	L	

AGG	AAT	GAA	ATT	CAA	TGG	TGA	ACT	ATC	AGA	TCT	ACA	AGA	CAA	AAT	TGI	TGG	AAT	TTT	GTC	9406
G	М	K	F	N	G	Ε	L	S	D	L	Q	D	K	I	V	G	Ι	L	S	
ACG	CCA	GGA	GCA	GGA	CTG	GCI	TTC	CAA	TGT	'ATA	AGT	ACA	тст	ТАТ	АСТ	CTC	AAT	AAA	TTG	9346
R	Q	Е	Q	D	W	L	S	N	V	*										
TTC	CAT	GTT	'AA'I	CTC	GTG	GAA	TGI	CAG	GGGG	ССТ	CGG	AGC	ATG	GCC	TAA	AAG	AAA	TGT	тст	9286
CAA	AAA	GTT	ACT	ACT	CCT	TCA	TGA	CCC	CAT	GAT	AGT	ATT	CAT	CCA	AGA	ATC	CAA	ACT	GGA	9226
ATG	TAT	TCC	TTC	TAA	ATT	GCA	AAA	ATC	'AAT	TTG	GTG	TGA	TGA	TGA	CCI	CAG	ССТ	CTG	TAT	9166
CAG	TCC	ATC	AAA	CGG	ATC	СТС	TGG	GAGG	ATT	'AAT	СТС	ССІ	ATG	GAG	ACC	CTC	AAA	ATT	TCA	9106
TCT	GGT	TTC	CAG	TAG	AAT	CGA	ATC	ACA	ATG	GAT	CGC	AAT	'GGA	AGG	AAT	GGT	GGT	GAG	GGA	9046
												М	Е	G	М	V	v	R	Е	
												>0	RF-	В						
AAA	TTT	TCA	ATG	CCI	тст	CAT	AAA	TAT	TTA	ТАА	СТС	СТС	TGA	TGC	TTC	GAC	TAG	ATC	AGA	8986
N	F	Q	С	$\mathbf{L}$	L	I	N	I	Y	N	s	С	D	А	s	т	R	S	D	
End	lonu	cle	ase	e/ex	onu	cle	ease	e/ph	losp	hat	ase	fa	mil	у (	pfa	um a	cce	ssi	on P	F03372)
CAC	ATG	GAA	CCA	TAT	'AGA	GGA	TTT	TTG	CAG	AAA	СТС	ACA	CTT	ACC	тст	TCT	AAT	AGC	GGG	8926
Т	W	N	Н	I	Е	D	F	С	R	N	s	Н	L	Р	L	L	I	А	G	
GGA	TTT	CAA	TGA	GGI	ACT	ATC	TTC	CCA	AGA	TCG	AGG	CAG	CCG	GAT	AAT	'AGA	TGA	AAC	TAG	8866
D	F	N	Е	V	L	S	S	Q	D	R	G	S	R	I	I	D	Е	Т	S	
TGC	CGG	AAA	ATT	CAG	GCA	ATT	CAT	AAC	CAA	ССТ	TCA	TCI	TAC	TGA	AAT	CAC	ACC	СТС	CAA	8806
А	G	K	F	R	Q	F	I	Т	N	L	Н	L	Т	Е	I	Т	Р	S	N	
TGG	АТА	CTT	TAC	ATG	GTT	TAG	GGGG	TCA	ATC	AAA	ATC	AAA	GCT	GGA	TAG	SAAT	тст	TGT	CCA	8746
G	Y	F	Т	W	F	R	G	Q	S	K	S	K	L	D	R	I	L	V	Q	
ACC	AGA	TTG	GAT	TCT	'GAA	ATT	CTC	ATT	CCI	TAA	TGC	СТС	CAT	ССТ	CAA	AAG	GAG	TAT	CTC	8686
Р	D	W	I	L	K	F	S	F	L	N	A	S	I	L	K	R	S	I	S	
L																				
GGA	TCA	TTG	CCC	TCT	TGT	'ACI	GAA	GTC	GCA	ATC	ТАА	GGA	CCG	GGG	ACC	GAA	ACC	TTT	CAG	8626
D	н	С	P	T.	v	T.	к	s	0	s	к	D	R	G	Р	к	P	F	R	
			-		•			5	£	5		2		U	-		-	-		
<u>አ</u> ጥጥ	יידירייד	тса	Сат	יכייר	CCT	יראר		CAA	CCA	ጥጥር	CCT	САЛ	ירריי	መልሮ	ሞልር		ልርጥ	ልጥር	CCA	8566
лтт Г	T.	т <u>о</u> д	м	W	T.	- С- Л-С - Т	ц	v v	ת ח	с С	T	т	т.	т Т	P	v v	17	W	F	0000
г	ц	U	11	vv	ц	т	п	л	U	C	ш	т	ц	т	л	А	v	vv	ы	
207	ጥጥር	GAA	GGG	: አ ጥ ጥ	יראר	יאמי	ካጥጥ ፖ	יםמי	GAA	ርጥጥ	ጥልእ	AGO	ጣርጣ	CAC	ת ב ב	ACA	ርጥጥ	GAN	ልርሞ	8506
AGA	rc	GAA	000		CAU	.AA1 -		AGA.	GAA	GTT F		AGC 7	,191			AGA	GTT T	GAA	AGT.	0000
D	5	ĸ	G	Ľ.	Т	T	5	E	ĸ	Ľ.	ĸ	А	v	к	ĸ	E	Ц	ĸ	v	

ATC	GAA	CCA	AGTO	CAAA	ATT	TGG	GAA	TCI	AGA	AAC	CAA	TAT	СТС	TCA	ATT	'AGA	AGA	CGA	AAT	84
W	N	Q	S	K	F	G	N	L	Ε	Т	N	I	S	Q	L	Ε	D	Е	I	
TCA	ACAA	ATG	GGG	ATAC	CTGI	TGC	CAA	CAC	GAG	AAA	CCT	ATC	GGT	TGA	TGA	ACI	GAG	TCT	CAG	83
Н	K	W	D	т	v	A	N	т	R	N	L	s	v	D	Е	L	S	L	R	
			-	-	•			-			_		•	-	-	_	-	_		
ATC	CCAA	GGC	CCL	ACI	IGGA	TTT	ATO	GGGA	ATTG	GAT	CAA	ACG	CAA	GGA	AAT	TCA	CTG	GGC	CCA	83
S	K	А	Q	L	D	$\mathbf{L}$	W	D	W	I	K	R	K	Е	I	Н	W	А	Q	
GAA	ACTC	TCG	TAT	raac	GTTG	GTT	GAA	GTG	TGG	GGA	TAA	GAA	СТС	AAA	GTI	CTT	CCA	TGC	СТА	82
N	s	R	I	s	W	L	K	С	G	D	K	N	s	K	F	F	Н	А	Y	
TGC	CATC	GAT	TAC	GAAG	SACG	GAA	GAA	TAA	TAT	CTC	TTC	CAT	CAC	GAT	CGA	TGG	TGA	GAC	CGT	82
А	S	I	R	R	R	K	N	N	I	s	s	I	т	I	D	G	Е	т	v	
СТС	GTGA	CCC	CGGF	AAA	AAI	CAA	AGC	CGA	AGC	CTC	ACT	СТА	TTT	CCA	AAA	TCI	GTT	CTC	AGA	8
С	D	Ρ	Е	K	I	K	А	Е	А	s	L	Y	F	Q	N	L	F	s	Е	
AGA	AAC	CTT	TTC	CCAG	GACC	AAC	TTT	CTT	GAA	ССТ	AGC	CTT	CAA	AAA	ACT	стс	ATC	AAT	ACA	8
Е	т	F	s	R	Р	т	F	L	N	L	А	F	K	K	$\mathbf{L}$	s	s	I	Q	
ATC	стс	GGA	ACCI	CAC	CAA	ACC	TTT	сто	ACA	СТС	TGA	AAT	AGA	AAA	AGC	AGI	AGC	ATC	ATG	8
S	S	D	L	т	K	Р	F	s	Н	S	Е	I	Е	K	А	v	А	S	С	
ТАС	GCCC	TTC	AAA	ATC	ccc	TGG	ccc	GGA	TGG	TTT	CAA	TTT	таа	СТТ	'CAT	AAA	GTC	TTC	CTG	7
S	Р	s	К	S	Р	G	Р	D	G	F	N	F	N	F	I	К	s	s	W	
-	-	-		~	-	-	-	2	-	-		-		-	-		-	2		
GGC	таат	CAT	CAP	AGZ	AGA	САТ	ካጥጥባ	יርͲር	аст	יידיG	CAA	тGA	ልጥጥ	СТС	GCA	GTC	'ፐGG	AAC	АСТ	7
Δ	Т	т	ĸ	E	Л	т	F	S	т.	v	N	E	F	W	0	s	G	 ጥ	т.	
11	-	1	к	Ц	D	T	1	D	Ц	v	I	Ц	1		Ą	U	U	1	Ц	
700	יא אר	CCC	ደመእር	ነ በ በ	ישכיי	יאכר	יכייים	יריסי	יאכר	CCT	יכאיד		C 7 7	сст		ACC	ירכר	ירייר	אאא	7
ACC	DAAG	000 C	C	JIAF M	11.01	AGC	נדט. ד	T T	.AGC	.GСI т	GAI			1001	GGA F	AGC	.ссс п	cic c	MAAA	7
Р	ĸ	G	5	IN	v	А	г	T	А	Ц	T	A	г	v	Е	А	Р	ъ	IN	
a=-		aa-			17.00				1000		<b>m a c</b>	000		a	<b>a - -</b>				0.000	-
CTT	'CAA	GGA	ACT'I	LCCC	SACC	CA'I	'CAG	F'A'I	'GG'I	'CGG	TAG	CCT	'I''I'A	CAA	GA'I	'AA'I	"TGC	GAA	GTT	1
F	K	D	F	R	Р	I	S	М	V	G	S	Ĺ	Ŷ	K	I	I	A	K	$\mathbf{L}$	
RNA	∖-de	pen	lder	nt I	DNA	pol	.yme	eras	se (	pfa	m a	cce	ssi	on	PF0	007	8)			
GCI	TTTC	CTT	CAC	GCI	GAA	AAA	TGI	'TAT	'GAA	CGA	тст	TAT	TGG	GCC	CCA	ACA	ATC	TTC	TTT	7
$\mathbf{L}$	S	F	R	$\mathbf{L}$	Κ	Ν	V	М	Ν	D	$\mathbf{L}$	I	G	Ρ	Q	Q	S	S	F	

TAT	TGA	GGG	GCG	CCA	GAT	СТТ	GGA	TAG	TGT	TTT	AAT	CAC	TGG	CGA	GTT	ATT	GGA	СТС	АТА
Ι	Е	G	R	Q	I	L	D	S	V	$\mathbf{L}$	I	Т	G	Е	L	L	D	S	Y
CAA	AAG	TTC	CAA	GAT	GGG	GGC	AGT	'AAT	GTT	AAA	АСТ	GGA	СТТ	CCA	CAA	GGC	СТТ	TGA	CAG
K	S	S	K	М	G	А	V	М	L	K	L	D	F	Η	K	Α	F	D	S
TGT	TTC	СТG	GTC	TTT	СТТ	GGA	TTG	GAC	CAT	GGA	TCA	ААТ	GGG	СТТ	CCC	ATT	AAC	ATG	GCG
V	S	W	S	F	$\mathbf{L}$	D	W	Т	М	D	Q	М	G	F	Ρ	L	Т	W	R
AAA	ATG	GAT	СТС	СТС	СТG	TGT	СТС	ATC	TGC	AGC	CGC	ATC	TGT	ССТ	CCT	AAA	TGG	СТС	TCC
K	W	Ι	S	S	С	V	S	S	А	А	А	S	V	L	$\mathbf{L}$	N	G	S	Р
TTC	GAC	тсс	GTT	CAA	GCT	CCA	GAG	GGG	ССТ	CCG	тса	AGG	AGA	CCC	TCT	СТС	TCC	СТТ	тст
S	Т	Р	F	K	L	Q	R	G	L	R	Q	G	D	Р	L	S	Р	F	$\mathbf{L}$
СТТ	TGT	GTT	AGC	AGC	GGA	AGT	TTT	GAA	тст	CAT	GAT	CAG	AAA	AGC	CAC	AGA	ATT	GAA	ТАА
F	V	L	А	А	Е	V	L	N	L	М	I	R	K	А	Т	Е	L	N	K
ATG	GTC	AGG	TAT	TGC	ТАТ	TTG	TAA	ATC	GGG	TCC	ТАТ	тст	AAC	тса	тст	TCA	ATT	TGC	AGA
W	S	G	I	A	I	С	K	S	G	Р	I	L	Т	Н	L	0	F	A	D
TGA	TAC	GAT	AGT	ATT	СТС	AAC	TCC	GGA	TTT	GAA	GGC	GCT	CAA	TAA	CAT	CCA	TAA	AAC	тст
D	т	I	v	F	S	т	Р	D	L	K	A	L	N	N	I	Н	K	т	L
САТ	CCT	GTT	CCA	GCT	ATC	стс	AGG	СТТ	GCA	GAT	CAA	СТТ	ССА	CAA	AAG	TGA	GAT	ССТ	TGG
Т	T,	F	0	T,	S	S	G	T.	0	Т	N	F	н	К	S	E	Т	T.	G
	_	-	z	_	-	-	•	_	z	_	_,	-			~	_	-	_	
ААТ		тас	TCC	тса	АТС	ጥጥG	GCT	ממידי	AGA	AGC	GGC	AAG	GCA	ልጥጥ	ידידאי	'ጥጥG	CAG	AGT	тGG
Т	N	т Т	P	0	S	W	T.	ĸ	E	Δ	<u>۵۵۵</u>	R	0	T.	F	<u>с</u>	R	v	G
		-	-	¥	0							1	¥		-	U		•	0
መ አ አ	ጥጥጥ	ccc	CAT		ሮሞአ	CCT	ccc	ССТ	ישפפ	አ አ ጥ	NCC	тсc	CAC	መመሮ	CCC	CAC	አጥጥ	ACC	7 7 C
			GAI				C	T			AGG	C	CAG	-11C		GAG	T	AGC	мас
IN	г	Р	T	Т	ĭ	Ц	G	Ц	Р	T	G	G	ъ	ъ	А	ĸ	Ц	А	Т
			<b>— — — —</b>		a a -	<b>a</b> -			a	a			<b>a</b>		a		a		
A'TG	igga	ACC	тст	CTT	GGA	GAG	AAT	GAG	GAA	GAA	ATT	GGC	CAC	A'TG	GAA	AGA	GAA	ATT	ACT
W	Ε	Ρ	$\mathbf{L}$	$\mathbf{L}$	Ε	R	М	R	K	K	L	А	т	W	K	Ε	K	L	L
СТС	GAT	TGG	TGG	AAG	ACT	CAC	TTT	ACT	AAA	AGC	СТС	АСТ	СТС	GAA	CCT	GCC	AAT	СТА	$\mathbf{T}\mathbf{T}\mathbf{T}$
S	I	G	G	R	$\mathbf{L}$	т	L	$\mathbf{L}$	Κ	А	S	L	S	Ν	$\mathbf{L}$	Р	I	Y	F

CAT	GTC	CTT	ATA	TCC	TAT	GCC	ACA	AGG	SAGT	TAT	'AGA	AAA	AAT	TAA	TAA	AAT	TCA	GAG	AAG	6886
М	S	L	Y	Ρ	М	Ρ	Q	G	V	I	Ε	K	I	N	K	Ι	Q	R	S	
СТТ	тст	TTG	GAG	TGG	TGG	TAT	GGA	TAA	AAG	GGC	TCT	ATC	TAT	GGI	'GAA	GTG	GGA	АТА	TGT	6826
F	L	W	S	G	G	М	D	K	R	A	L	S	М	V	K	W	Е	Y	v	
CCA	GCT	тсс	AAA	AGC	GTT	GGG	AGG	CTI	AAA	TGT	'GAG	TAA	ССТ	TCI	GAT	TAG	AAA	TTT	GGG	6766
Q	L	Ρ	K	A	L	G	G	L	N	V	S	N	L	L	I	R	N	L	G	
GCI	CCT	TTG	TAA	GTG	GGI	GTG	GAG	GTA	TTT	TTC	AGA	ACC	AGA	TTC	GCT	ATG	GAG	ACT	ATC	6706
L	L	С	K	W	V	W	R	Y	F	S	Е	Ρ	D	S	L	W	R	L	S	
AAT	TAA	AGC	CAA	ATA	TAA	АТА	CCA	GGC	GCA	AAT	'GAA	TAT	GGC	TGA	CAT	TGC	TCC	AAT	AAG	6646
I	K	A	K	Y	K	Y	Q	A	Q	М	N	М	A	D	I	A	Ρ	I	R	
ATC	AGG	TGG	TCC	TTG	GAG	ACA	тст	TTG	GCAA	CCA	тст	ССТ	AAA	ACA	CCA	AGC	AAC	AAA	TGA	6586
S	G	G	Ρ	W	R	Н	L	С	N	Η	L	L	K	Η	Q	A	Т	N	Е	
ACT	TCT	GAA	ACA	AGG	тас	CAG	GAA	AAG	SAAT	'AGG	GAA	TGG	GGA	GAA	TAC	CTI	ATT	TTG	GCA	6526
L	L	K	Q	G	Т	R	K	R	I	G	N	G	Ε	N	т	L	F	W	Н	
TGA	СТС	TTG	GCI	GGG	CAA	TTT	GCC	TCT	GAA	ATT	AAC	CTT	ccc	AAG	ACT	CTT	CCT	'AAT	CTC	6466
D	S	W	L	G	N	L	Ρ	L	K	L	Т	F	Ρ	R	L	F	L	I	S	
AGT	TTT	ACC	CAI	GGC	TTC	AGT	'AGC	GGA	GAT	GGG	TTC	TTG	GGT	GAA	TTT	GGA	ATG	GAA	ATG	6406
V	L	Ρ	М	A	S	V	A	Е	М	G	S	W	V	N	L	Е	W	K	W	
GAA	TTT	GCC	ATG	GTC	CAG	AGA	ATT	CAG	SAAA	GAG	AGA	CCG	CAT	TGA	ATG	GGA	GCA	GCT	CCA	6346
N	L	Ρ	W	S	R	Е	F	R	K	R	D	R	I	Е	W	Е	Q	L	Q	
ACC	TTC	сст	CCA	GCA	AAT	CTC	AGT	000	сст	CAA	TGA	ATC	AGA	TGA	GTT	AAT	ATG	GAA	CTT	6286
Ρ	S	L	Q	Q	I	S	V	R	L	N	Ε	S	D	Е	L	I	W	N	F	
TAG	TAT	GGC	TGG	TAA	TTT	CTC	AGT	TCG	стс	CTT	СТА	TGA	AGA	ACI	TCA	CAA	GCG	CTC	GAA	6226
S	М	A	G	N	F	S	v	R	S	F	Y	Ε	Ε	L	Н	K	R	S	K	
GCC	CTG	TCT	'AGA	AAA	тст	CCC	TCA	AAA	GAT	ATG	GAA	AGG	ACT	TGI	TCC	CTT	CCG	AAT	AGA	6166
Р	С	$\mathbf{L}$	Е	N	$\mathbf{L}$	Ρ	Q	K	I	W	K	G	L	v	Ρ	F	R	I	Е	

AAT	CTT	CAC	TTG	GTI	ATC	AGT	GCI	AGA	GAG	AAT	CAA	TAC	TAA	GAA	GAA	ACI	AGC	TTC	тст	6106
I	F	Т	W	L	S	V	L	Ε	R	I	N	Т	K	K	K	L	A	S	L	
GAA	CAT	TAT	CCC	ACC	CGC	TGA	GGT	GGG	TTG	CTC	ATT	ATG	TAG	TTT	GGA	GCC	TGA	GGA	TAT	6046
N	I	I	Ρ	Ρ	A	Е	v	G	С	S	L	С	S	L	Е	Ρ	Е	D	I	
TTC	GCA	CCI	CTT	TTT	GTT	TTG	CCC	CTI	стс	AAT	'GGA	GAT	TTG	GGC	TTG	GTG	GTG	GGA	ССТ	5986
S	Η	L	F	L	F	С	Ρ	F	S	М	Ε	I	W	A	W	W	W	D	L	
ТТG	GAA	CCI	'ATC	TTG	GGI	ATG	GCC	AAA	ATC	TCT	AAA	тст	TGC	сст	СТС	TCA	ATG	GAA	TTG	5926
W	N	L	S	W	V	W	Ρ	K	S	L	N	L	A	L	S	Q	W	N	С	
ccc	AAG	GAA	GGA	AAA	ATT	ATT	CAA	AAA	AAT	CTG	GCT	GGC	AGC	ATT	CAT	TGT	GAT	TAT	CTG	5866
Ρ	R	K	Ε	K	L	F	K	K	I	W	L	A	A	F	I	V	I	I	W	
GTC	AAT	CTG	GAG	AGA	ACG	CAA	TGA	GAG	SAAT	TTT	CAA	TAA	GAA	AGA	ATC	ATC	AGT	TTC	AGA	5806
S	Ι	W	R	Е	R	N	Ε	R	I	F	N	K	K	Ε	S	S	V	S	Е	
AAT	'CAA	AAA	CCI	CAT	TCT	TGT	CCG	TTT	ATG	TTG	GTG	GAT	GAA	GCC	TTG	GAA	CCI	CTC	СТТ	5746
I	K	N	L	I	L	v	R	L	С	W	W	М	K	Ρ	W	N	L	S	F	
ссс	GTA	CAC	AAT	TGA	AGA	AGT	CAT	CAG	SAAT	CCC	ACA	ATG	тст	СТТ	ATG	GGG	TAG	CGC	TGT	5686
Ρ	Y	Т	I	Е	Ε	V	I	R	I	Ρ	Q	С	L	L	W	G	S	A	v	
GCC	TCG	AAG	AAG	TAA	AAC	стс	CCA	тст	ccc	ccc	тст	AAT	тса	GCT	CAG	ATC	TAA	ccc	ссс	5626
Ρ	R	R	S	K	Т	S	Н	L	Ρ	Ρ	L	I	Q	L	R	S	N	Ρ	Ρ	
TGA	ccc	TTG	TCI	CAA	GTG	GAT	GGI	GGG	TTT	CAC	ccc	GTT	СТС	GCC	AAA	AGA	AGG	TGC	TAG	5566
D	Ρ	С	L	K	W	М	V	G	F	Т	Ρ	F	S	Ρ	K	Е	G	A	R	
AGC	AGG	AGG	CAI	TTT	TGG	AGG	СТТ	rcci	CAG	AGA	TGA	AGT	GGG	TGT	GAT	CTT	ATG	СТС	СТТ	5506
A	G	G	I	F	G	G	F	L	R	D	Е	V	G	V	I	L	С	S	F	
СТС	CTG	CCC	TTT	TCC	GCC	AAT	GGG	TAT	TAA	TGA	AGT	TGC	AGT	GAT	TGC	AAT	TCA	CCG	AGC	5446
S	С	Ρ	F	Ρ	Ρ	М	G	I	N	Е	V	A	V	I	A	I	Η	R	A	
тст	'GCA	AAT	СТС	TCT	CAG	TGT	GCA	AAA	TCT	'AAA	AGA	CCG	AGA	AAT	СТС	AAT	TTT	CTC	TGA	5386
L	Q	I	s	$\mathbf{L}$	S	v	Q	N	L	K	D	R	Е	I	s	I	F	s	Е	

ATC	CAG	CCA	AGC	TAT	CAG	TTG	GTG	CCT	CAA	тст	CTC	ATC	TGG	TCC	GAC	TAA	тст	CTC	CTT	5326
S	S	Q	A	I	S	W	С	L	N	L	S	S	G	Ρ	т	N	L	S	F	
ССТ	GTT	GAA	СТТ	CAT	CAG	АТС	TAC	ATG	CAA	AAA	GCT	ccc	TCT	тст	GAA	GTT	TGA	TTA	тст	5266
L	L	N	F	I	R	S	Т	С	K	K	L	Ρ	L	L	K	F	D	Y	L	
ርጥር	AAG	ርጥር	ርጥሮ	AAG	ምሮል	ልርጥ	מממ	ъсъ	GDD	ጥርር	ית מי	TCC	ACA	ልልጥ	ርጥል	ጥርጥ	ՠՠՠ	CTTC	AGA	5206
010	1110	010		1110	1011	101	11111	11011	101111	100	11111	100	11011	11111	CIN	TOT	<b>T T T</b>	010		5200
S	S	С	S	S	Q	V	K	Q	N	A	Ι	G	Ε	Ι	Y	V	F	S	D	
TGT	AGT	TAG	ATG	GAA	AAA	GTC	TCC	CAT	TTA	ATT	TTG	AAC	CGA	ccc	ССТ	ccc	ATG	GTA	TGA	5146
v	V	R	W	K	K	S	Ρ	I	*											
TGT	AAG	TGG	CAG	ACG	AAA	GTA	CCC	TTT	TAA	AAG	GAT	TGT	'GAA	АТА	AAA	TGA	TAA	AAA	AAA	5086
AAA	AAC	ACA	САТ	CAT	AAG	ТАА	CCG	ATT	AAA	CAA	GTC	ACA	AAT	AGG	тта	AAA	СТС	AAT	'AGA	5026
AGA	AAA	CAG	ACA	TAC	AAA	ССТ	AAC	тса	ACC	AAG	GAT	AAT	CAA	AAG	GTA	GAT	TTA	AAA	TGA	4966
GAA	AAC	GCG	САА	ААА	ттG	АСТ	тат	тта	AAA	CAA	ידיד	TTC	CAA	ААА	ААА	ААТ	GAT	АТА	AAC	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.

TGAGTTTTTCATACTTCAGAACAGTGTGTAATATCCCCAAATTTTTATAACTATTTTATAA 79114 ΑΑΤΤΑΑΤΤΤΤΤGTTATAATAAGATTTTATATATATATATATTCTGAAAAATAAAGTAAAATC 79054 GAAATTTCGAAATCAAAATTTTGAAATTAAATCAGATTTTTATCTTTTGAAGAAAAAAA 78934 AATTGGAATTTGATTTTCAGATTTTTCGTCCAAAACGAAAAATAGAGAGAAAAGAAAATT 78874 CTGAAATTATAATTTGAGTTGGTTTTGGAAAAGGATTAGATTTTTGTAAATACTTATCTT 78814 CACGTAATACACTTTCTTCATCTTTTTGGTAAGTTTTAAATTTCAGATCTAAAATCACCA 78694 TTGTTGTGTTTGAAGGTTTCAACAAAAAAAAAAAAAACTTTTAAAAATCCGATGACTTGGT 78634 GCGTGCGATGCTTCACAGTACCTCCAAGATTTGATGTTTTTAAAATAAAAAAAGAAAAAA 78454 AAAAAAAGAAAAAGGAATTACATGTGCATTAAAGTCCACGAGAGATTAGACTTACTGTAA 78394 CATGAAAGTTGAAAAAAATTTCAAGAAAAAATTTCCTTTCAATCTGTGACGACGCTGTC 78334 ATATTAAGTGACTAACATTAGGTAATATTATTTGTTGTAGACGGAAACTTCGTAGAAGGC 78154 GATTTTTAGTTGATATTGCTATACTTGGGACGCGTTTGAGGTAATTCACAATGTCCATCA 78094 ACAACTAGAATCCCGTTAGGAATTGTATGTGCCTATGTGTAATGCATGTGGTTATGTGATT 78034 CTATATTTATGTTAGATGTATATTATGTGACAGGAATCGTATAGTACTTTCTATGACTAA 77974 ACTGGTATTGTTATGAATTTGGATTTATAATGTACCTAATTATACCATCGTATTAAA 77854 ΑΤΤΑΤGCTAATATTGACAATGTACTTAATGGGTATAACTAGTGTTGATGATGTAGCGA 77794 ACGGTATTATAAGTTGACATATATTGTAACTCTATGAGGCTCTAATGATGGATATATTGG 77734 TTACTACGAATTATGGACGTTGATTAGTATTAAGTGGCTAAGTTGTGAAAAATATTATTT 77674 GAACTAAAGATGTTCCTGCTAATGTTAATGTGATGTTTGATGTGTCACAACTTTTAAAAA 77614 TCTATTAATCACGTAAAGTGGAAATTGAGGGAATTATCCTGTGGATTTGGATCCTCCATA 77554 GGTGATGAACAGTACTTGATTTTATTATGATACAACTCTTTATTGTCTTCCTCCTAATAC 77494 TATTGGTGCGCATTGCGGATACCCATTAGATTAGTGAGGGGTGTGCACACTAGGGACGCA 77434 CTGCAAAACGATATTGGCCATTGCTCTTAATTTATTGGGTGACGACCCATGTTGAAGTAG 77374 GTGGTTACTGGGATTAGTCCCGGCCTACTGACGTTCGATTCCCCTCTAATAATTAAATGT 77314 GACGTTGTGTGCATCATATAAACTTGGAGTATATAAATTGATTAAGTGTGTTATGTACTT 77254 ATTTTTAAAATACAACTATTATTATTATTATTATTACGGGATGGAGTTGTAATTACTTAGC 77134 TTTCGCTAATTTTTGTGTTTTTGTTTTTGTTTTGCTCTTTTCCTATTTATATTGTGCAGGT 77074 TGGTGAAAGGGACTACGTTGCAGGAATGAGTGAATTTATAGTTTTTAAGTTCACCTAGGT 77014 TAATAAACACTACAGTACTGCCATTTGAGATTGTTTAGTTTTGGTATAAAGTAGA 76954 CAACTCAGTTTGGTTTATTTTGATATTTGGTTTGGGTTGGGGTAAAATTTTTATTTCTAATAA 76894 TTAGCCTTGCATTTATTTTGAACAATAGATGCGGCGGTAACACCCTAAAGGTTTGGTCAG 76834 ATATTTTAATGAAAAGTGAATATTTTATAAAGTTTTAGAACCAAAGTTTTGGGGTGTTAC 76774

TGTTAAATAATTAAGAGTCTAAGATTAAATTAAGTAATAAGCATTAAAATAAAT	AAAT	TGG	TAT	TCA	GAG	СТТ	AGG	TTF	ATAA	TAA	ATA	AAT	ATA	LAAI	TAG	GAG	TTA	АТА	AAA	TT	76714
AGTGGGATCGANAAGGGTTTATGTGCCTTANTGTGAAAGCACTTATTTCAGGTGCTTATT 7654 AATTGTTTCTTTCGTTCTGTATGTGCCTTAATGTGAAAGCACTTATATCACAGAGAAAATGTTGTGCCC 76534 GGATGACCGACCATAGGATTTAGAGATTTATAGAATGTTGTATATAGGACGGTACTAAGGATTTAGAGTTTTATGTGTAAGGTTATATGATGATG	TGTI	AAA	TAA	TTA	AGA	GTC	ТАА	GAI	TAA	ATT	AAG	TAA	TAA	GCA	ATTA	AAA	TAA	ATT	GTG	AG	76654
AATTGTTCTTTCGTTCTTGTATGTTAGATGTTGTAACTACAAGAAAATGTTGTGCC 76534 GGATGACCGACCATAGGATTAGAGATTTAGAGATTTATATATA	AGTO	GGGA	TCG	AAA	AGG	GTT	TAT	GTC	GCCT	ТАА	TGT	GAA	AGC	ACI	TAT	TTC	AGG	TGC	TTA	TT	76594
GATGACCGACCATAGGATTTAGAGATTTATATTATATTA	AATI	GTT	TCT	TTT	CGT	тст	TGT	ATC	GTTT	AGA	TGT	TGI	'AAC	TAC	CAAG	AAA	ATG	TTG	TGC	тС	76534
AGTATTGAAAATTCTCTAAGTTCCTATTTATGTGTTTTAATTATGAATCTTATTACTTGA 76414 CTATTTGAGTTGATTAGAGTTTAAGTTCTTGTTATAATTATTGCTTACTTGCCAAATTGA 76354 TGTTATGTGGAAGCATCCAAAGGTTGCTTGTAAATTATGCTTACTGCAAATTAAGTTG 76294 GTTGAACTACGAAATTTTCTTATGATGAAAGTTGTATGTA	GGAI	GAC	CGA	CCA	TAG	GAT	TTA	GAG	GATT	TAT	ATT	AAG	GTC	CGI	ATA	CGA	AGT	TAT	TTA	GT	76474
CTATTTGAGTTGATTAGAGTTTAAGTTCTTGTTATAATTATTGCTTACTTGCTAATTGGAAATTGAT GTTATGTGGAAGCATGCAAAGGTTGTCTTGAAATGTATGT	AGTA	ATTG	AAA	ATT	CTC	TAA	GTT	CCJ	TTAT	ТАТ	GTG	TTT	TAA	TTA	TGA	ATC	TTA	тта	CTT	GA	76414
TGTTATGTGGAAGCATGCAAAGGTTGTTTGAAATGTATGT	CTAI	TTG	AGT	TGA	TTA	GAG	TTT	AAG	GTTC	TTG	TTA	TAA	TTA	TTO	SCTT	ACT	TGT	CAA	ATT	GA	76354
GTTGAGTCATGAAATTTTCTTATGATGAAAGTTTGATTGA	TGTI	ATG	TGG	AAG	CAT	'GCA	AAG	GTI	[GTC	TTG	AAA	TGI	ATG	TTA	TGA	TGA	AAT	АТА	GTT	ΤG	76294
AGACCATGTATGGGTGAATAGATGTAGTAGTAGTAGTAGCAATGCGAGGAGTGCAGGAATTGTATGC       76174         M       L       M       C       R       N       C       M       Q         AAACCTATCAAGTATACACTAGATCTCGTAGGAGGACTTTTGAACAATGGCTGAAACCA       7       R       S       R       R       T       F       E       Q       M       A       E       T       P         CTGAACAACTACTGGAGAATTGAGAATGAGATCTCTTGAACAACTTCCCAACGTATGGGTATGGGTTAGGAACAACTACTGGGAAACCAACTACTGGGAAACCAACTTGGGAAACCAACGTGGTAGGAAACCAACGTGGGAGAACCAACGACACGGCTAGGGCAAAGA       76054       E       Q       N       A       E       T       P       76054         L       Q       L       E       R       L       R       S       L       E       Q       L       S       Q       R       M       G       L       V         CTGTACAAAACCAATTAGGAAATAATGGTGGAGAGGAGGAGGACCCACAAGGCTGCAAAGGCAACCTTAGGCAAACCCTTAAACTTCAAAGTTATCTGTAGGAAGGA	GTTO	GAGT	CAT	GAA	ATT	TTC	TTA	TGF	ATGA	AAG	TTT	GTA	TTG	AGI	'AAA	TAA	AAC	GAT	GTG	CA	76234
AGACCATGTATGGGTGAATAGATGTAGTAGTAGTAGTAGCAATGGCTGAGTGGGAATGGTAGCTAGGGTGGGAGATGGGGAGTGCCAAGGGAGGACCTTTGAACAATGGCTGGAAACCC AAACCTACTAGGAGAGTGAGATGAGATCTCGGAGGGGCCTTTGAACAATGGCTGGAAACCC TGTACAAAACCAACTACTGGGAGAGAATAATGGTGGAGAGGACCCACAAGCTGCTATGGCAAAGA CTGGTACAAAACCAATTAGGAAATAATGGTGGAGAGGAGCCCACAAGCTGCTATGGCAAAGA C Q L L E R L R S L E Q L S Q R M G L V TGTTACAAAACCAATTAGGAAATAATGGTGGAGAGGACCCACAAGCTGCTATGGCAAAGA TGTTACAAAACCAATTAGGAAATAATGGTGGAGAGGACCCCACAAGCTGCTATGGCAAAGA C Q N Q L G N N G G E D P Q A A M A K K AGTTAGCAACCCTTTAAACCTCCAATCTTGTAGGAAAGGAAGACCCCCTTACTCTTAGAGA L A T L K P P I F V G K E D P L L L E N ACTGGCTAAGAGACTTTGATAAGTTATTCACTGCTACTGGGACACCTGAAGCTCAAAAG 75874 W L R D F D K L F T A T G T P E A Q K V TAGACCAAGCTACCTTTGTAGGGAAACTTAATGGAGGCAGACCTTGGTGGGAGAGTCAAGGAC 75814 D Q <u>A T F Y L R E D A D T W W E S Q G P</u> RetroTranspoon gag protein (pfam accession PF03732) CTATTGTTAGAGCCCAGGAAACTTTAATGGAAGCACACTTTGAGGTGCCAGAGTTAAGGTT ACTGGCTAAGAGACTTTATTGGAAACTTTAATGGAATGCTTTTAAGGTTCCTATTAACGATTTAACGAG 75694 <u>F F P E H I R Q K Y N E F T R F N Q G</u> GAGGTACTATGTCTGTGCAAGAGTACCCCAAAGGTCCAAGGATTGCCTAGATTTACCCAG G T M S V Q E Y A Q K F N E L A R F C P																					
M       L       M       C       R       N       C       M       Q         AAACCTATCAAGTATACACTAGATCTCGTAGGAGGACTTTTGAACAAATGGCTGAACAATGCTGAAACC       Y       Q       V       Y       T       R       S       R       R       T       F       E       Q       M       A       E       T       P         CTGAACAACTACTTGAGAAATAATGGAGAATGAGATCTCTTGAACAACTTCGCAACGTACGGATGGGATGGAGACCCACAAGGAACGAAC	AGAC	CAT	GTA	TGG	GTG	AAT	AGA	TGI	TAGT	ААТ	ACC	AAT	GCI	'GA'I	GTG	CAG	GAA	TTG	TAT	GC	76174
AAACCTATCAAGTATACACTAGATCTCGTAGGAGGAGGAGGAGGAGGACTTTTGAACAAATGGCTGAAAACC76114TYQVYTRSRRTFEQMAETPCTGGAACAACTACTTGGAGGATTGAGATTGAGATCTCTTGAACAACTTTCGCAACGTATGGGGTTAGG76054EQLLRRSLEQLSQRMGLVTGTTACAAAACCAATTAGGAAATAATGGTGGGAGGGCCCACAAGCTGCTATGGCAAAGA75994LQNQLGNNGGEDPQAAMAKKAGTTAGCAACCCTTAAACCTCCAATCTTTGTAGGAAAGGAAAGGAAGG												М	$\mathbf{L}$	М	С	R	N	С	М	Q	
AAACCTATCAAGTATACACTAGATCTCGTAGGAGGACTTTTGAACAAATGGCTGAAACCC       76114         T       Y       Q       V       Y       T       R       S       R       R       T       F       E       Q       M       A       E       T       P         CTGGAACAACTACTTGAGGAGTTGAGATCTCTTGAACAACTTTCGCAACGTTGGGAACGATTGGGGAAGGACCCACAAGGTGCTATGGCAAAGA       76054       E       Q       L       L       R       R       R       R       P       Q       R       M       G       L       V         TGTTACAAAACCAACTACTGGAGAATAATGGTGGAGAGGACCCACAAGGCTGCTATGGCAAAGA       75994       L       Q       N       Q       L       G       N       N       G       G       E       D       P       Q       A       M       A       K       K         AGTTACAAAACCAATATAGGAAATAATGGTGGAAGGACCCACAAGGAACCCTTACAAGACACCTTAGAAAACCAACTTTAAACCTCCAAAACTTTAGGAGGAAGAACCCCCACAAGGACCCTTAGAACTTTAAGAAC       75934       L       A       T       L       K       K       K       75934         L       A       T       L       K       P       P       I       F       V       G       K       E       D       P       L       L       L       N       76374         M      L																					
T       Y       Q       V       Y       T       R       S       R       R       T       F       E       Q       M       A       E       T       P         CTGGAACAACTACTTGAGAGATTGAGATCTCTTGAACAACTTTCGCAACGATGGGGTTAGGGGTTAGGGGACAACCAAC	AAAC	СТА	TCA	AGT	АТА	CAC	TAG	АТС	CTCG	TAG	GAG	GAC	TTT	TGA	ACA	AAT	GGC	TGA	AAC	CC	76114
CTGAACAACTACTTGAGAGATTGAGATCTCTTGAACAACTTTCGCAACGTTAGGGTTTAG       76054         E       Q       L       L       E       R       L       R       S       L       E       Q       L       S       Q       R       M       G       L       V         TGTTACAAACCAATTAGGAAATAATGGTGGAGGAGCCCCACAAGCTGCTATGCCAAGCAATGCAATGGAAATAATGGTGGAGGAGCCCCACAAGCTGCAAGGCAACCAAGGCAACCCTTAGAAAA       75994       L       Q       N       Q       L       G       N       N       G       G       E       D       P       Q       A       M       A       K       K         AGTTAGCAACCCTTTAAACCTCCAATCTTTGTAGGAAAGGAAAGGAAGG	Т	Y	Q	v	Y	т	R	s	R	R	R	т	F	Е	Q	М	А	Е	т	Ρ	
CTGAACAACTACTTGAGAGATTGAGATTGAGATTCTTGAACAACTTTCGCAACGTATGGGTATGGGTTTAG       76054         E       Q       L       L       R       L       R       S       L       E       Q       L       S       Q       R       M       G       L       V         TGTTACAAAACCAATTAGGAAATAATGGTGGAAGAGGACCCACAAGCTGCTATGGCAAAAC       75994       L       Q       N       Q       L       G       N       N       G       G       E       D       P       Q       A       M       A       K       K         AGTTAGCAACCCTTAAACCTCCAATCTTTGTAGGAAAGGAAGG																					
E       Q       L       L       R       R       S       L       E       Q       L       S       Q       R       M       G       L       V         TGTTACAAAACCAATTAGGAAATAATGGTGGAGAGGACCCACAAGCTGCTATGGCAAAAC       Q       N       Q       L       G       N       N       G       G       E       D       P       Q       A       A       M       A       K       K         AGTTAGCAACCCTTAAACCTCCAATCTTTGTAAGGAAAGGAAAGGAAGACCCCTTACTCTGGAAAGGACCCCTTAAGAGA       75934       L       A       T       L       K       P       P       I       F       V       G       K       E       D       P       L       L       L       E       N         ACTGGCTAAGAGACCTTTGATAAGTTATCACTGCTACGGGACACCTGGGACCCCGAAGGCCAAGGAGAGACCCTTTTAAGGAGACCCTTTTAAGGAGAGAGA	CTGA	ACA	ACT	ACT	TGA	GAG	ATT	GAG	GATC	тст	TGA	ACA	ACI	TTC	GCA	ACG	TAT	GGG	TTT	AG	76054
TGTTACAAAACCAATTAGGAAATAATGGTGGAGAGGACCCACAAGCTGCTATGGCAAGA       75994         L       Q       N       Q       L       G       N       N       G       G       E       D       P       Q       A       A       M       A       K       K         AGTTAGCAACCCCTTAAACCTCCAATCTTTGTAGGAAAGGAAGG	Е	Q	$\mathbf{L}$	$\mathbf{L}$	Е	R	L	R	s	L	Е	Q	$\mathbf{L}$	s	Q	R	М	G	L	v	
TGTTACAAAACCAATTAGGAAATAATGGTGGAGGAGCCACAAGCTGCTATGCCAAGA       75994         L       Q       N       Q       L       G       N       N       G       G       E       D       P       Q       A       A       M       A       K       K         AGTTAGCAACCCTTAAACCTCCAATCTTTGTAGGAAAGGAAGACCCCTTACTCTTAGGAA       A       M       A       K       K       75934         L       A       T       L       K       P       P       I       F       V       G       K       E       D       P       L       L       L       E       N         ACTGGCTAAGAGACTTTGATAAGTTATTCACTGCTACTGGGAACCCTGAAGCTCAAAAG       75874       M       L       R       D       F       D       K       L       F       T       A       T       G       T       P       E       A       Q       K       V         TAGACCAAGCTACCTTTTATTCTGAGGGAGGAGGAGGAGGAGCAGAGCACCTTGGAGGGGAGGAGCACCTTTTAAGGGAGCTAAGGAGAGACTTTAATTGGAGGAGGAGCACCTTGGAGGGAG																					
L       Q       N       Q       L       G       N       N       G       G       E       D       P       Q       A       A       M       A       K       K         AGTTAGCAACCCTTAAACCTCCAATCTTTGTAGGAAAGGAAGG	TGTI	ACA	AAA	CCA	ATT	'AGG	AAA	TAF	ATGG	TGG	AGA	GGA	CCC	ACA	AGC	TGC	TAT	GGC	AAA	GA	75994
AGTTAGCAACCCTTAAACCTCCAATCTTTGTAGGAAAGGAAGG	L	Q	N	Q	L	G	N	N	G	G	Е	D	Ρ	Q	А	А	М	А	K	к	
AGTTAGCAACCCTTAAACCTCCAATCTTTGTAGGAAAGGAAGCCCCTTACTCTTAGAGA       75934         L       A       T       L       K       P       P       I       F       V       G       K       E       D       P       L       L       L       E       N         ACTGGCTAAGAGACTTTGATAAGTTATTCACTGCTACTGGGACACCTGAAGCTCAAAAG       75874       M       L       R       D       F       D       K       L       F       T       A       T       G       T       P       E       A       Q       K       V         TAGACCAAGCTACCTTTTATCTGAGGGAGGAGGAGGAGGAGGAGAGAGA																					
L       A       T       L       K       P       P       I       F       V       G       K       E       D       P       L       L       L       E       N         ACTGGCTAAGAGACTTTGATAAGTTATTCACTGCTACTGGGACACCTGAAGCTCAAAGCTCAAAAAG       75874       W       L       R       D       F       D       K       L       F       T       A       T       G       T       P       E       A       Q       K       V         TAGACCAAGCTACCTTTTATCTGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	AGTI	AGC	AAC	ССТ	TAA	ACC	TCC	AAJ	CTT	TGT	'AGG	AAA	GGA	AGA	ACCC	СТТ	ACT	СТТ	AGA	GA	75934
ACTGGCTAAGAGACTTTGATAAGTTATTCACTGCTACTGGGACACCTGAAGCTCAAAAAG       75874         W       L       R       D       F       D       K       L       F       T       A       T       G       T       P       E       A       Q       K       V         TAGACCAAGCTACCTTTTATCTGAGGGAGGGAGGCAGGCA	L	А	т	$\mathbf{L}$	K	Р	Ρ	I	F	v	G	К	Е	D	Ρ	L	L	$\mathbf{L}$	Е	N	
ACTGGCTAAGAGACTTTGATAAGTTATTCACTGCTACTGGGACACCTGAAGCTCAAAAG       75874         W       L       R       D       F       D       K       L       F       T       A       T       G       T       P       E       A       Q       K       V         TAGACCAAGCTACCTTTTATCTGAGGGAGGAGGAGGAGGAGCACACTTGGTGGGAGAGTCAAGGAC       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)       75754         I       V       R       A       Q       E       N       F       N       N       A       F       K       V       A       I       K       D       R         GATTTTCCCTGAACATATTAGGAGGCAGAAATACAATGAGTTCACTAGATTTAACCAGG       75694       F       F       N       N       A       F       K       V       A       I       K       D       R         GATTTTCCCTGAACATATTAGGAGGCCCAGAAATACAATGAGTTCACTAGATTAGCTAGATTTGCC       75634       G       T       N       S       V       Q       K																					
W       L       R       D       F       D       K       L       F       T       A       T       G       T       P       E       A       Q       K       V         TAGACCAAGCTACCTTTTATCTGAGGGAGGATGCAGACACTTGGTGGGAGAGTCAAGGA       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)       75754         I       V       R       A       Q       E       N       F       N       N       A       F       K       V       A       I       K       D       R         GATTTTCCCTGAACATATTGGGAGGCAGAAAACTTTAATTGGAATACAATGAGTTCACTAGATTTAACGATTAACCAAGG       T       N       N       A       F       K       V       A       I       K       D       R         GATTTTCCCTGAACATATATGGGAGGCAGAAATACAATGAGTTCACTAGAGTTCACTAGATTAACCAAGG       F       T       R       F       N       N       E       F       T       R       N       Q       G       75634         GAGGTACTAT	ACTO	GCI	'AAG	AGA	CTT	'TGA	TAA	GTI	TTAT	CAC	TGC	тас	TGG	GAC	CACC	TGA	AGC	тса	AAA	AG	75874
TAGACCAAGCTACCTTTATCTGAGGGAGGAGGAGGAGGAGAGCACTTGGTGGGAGAGTCAAGGAC75814DQATFYLREDADTWWESQGFRetrotransposongagprotein(pfamaccessionPF03732)CTATTGTTAGAGCTCAGGAAAACTTTAATTGGAATGCTTTTAAGGTTGCTATTAAGGATA75754IVRAQENFNWNAFKDRGATTTTCCCTGAACATATTAGGAGGCAGAAAATACAATGAGTTCACTAGATTTAACCAGG75694FFPEHIRRQKYNEFTRFNQGGAGGTACTATGTCTGTGCAAGAGTATGCCCAAAAGTTCAATGAGTTCACTAGATTTAGCTAGATTTTGCC75634GTMSVQEYAQKFNELARFCP	W	L	R	D	F	D	K	L	F	т	A	т	G	т	Р	Е	A	Q	K	v	
TAGACCAAGCTACCTTTTATCTGAGGGAGGAGGAGGAGGAGAGGAGGAGGAGGAGGAGGAG																					
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) CTATTGTTAGAGCTCAGGAAAACTTTAATTGGAATGCTTTTAAGGTTGCTATTAAGGATA 75754 I V R A Q E N F N W N A F K V A I K D R GATTTTTCCCTGAACATATTAGGAGGCAGAAATACAATGAGTTCACTAGATTTAACCAGG 75694 F F P E H I R R Q K Y N E F T R F N Q G GAGGTACTATGTCTGTGCAAGAGTATGCCCAAAAGTTCAATGAGTTAGCTAGATTTGCC 75634 G T M S V Q E Y A Q K F N E L A R F C P	TAGA	ACCA	AGC	TAC	CTT	TTA	тст	GAG	GGGA	GGA	TGC	AGA	CAC	TTC	GTG	GGA	GAG	тса	AGG	AC	75814
Retrotransposon gag protein (pfam accession PF03732)CTATTGTTAGAGCTCAGGAAAACTTTAATTGGAATGCTTTTAAGGGTTGCTATTAAGGATAT N N N A F K V A I K D RGATTTTTCCCTGAACATATTAGGAGGCAGAAATACAATGAGTTCACTAGATTTAACCAGGGATTTTTCCCTGAACATATTAGGAGGCAGAAATACAATGAGTTCACTAGATTTAACCAGG75694F F P E H I R R Q K Y N E F T R F N Q GGAGGTACTATGTCTGTGCAAGAGTATGCCCAAAAGTTCAATGAGTTAGCTAGATTTGCC75634G T M S V Q E Y A Q K F N E L A R F C P	D	Q	А	Т	F	Y	L	R	Е	D	A	D	Т	W	W	Е	S	Q	G	Ρ	
CTATTGTTAGAGCTCAGGAAAACTTTAATTGGAATGCTTTTAAGGTTGCTATTAAGGATA75754IVRAQENFNWNAFKVAIKDRGATTTTTCCCTGAACATATTAGGAGGCAGAAATACAATGAGTTCACTAGATTTAACCAGGFFPEHIRRQKYNEFTRFNQGGAGGTACTATGTCTGTGCAAGAGTATGCCCAAAAGTTCAATGAGTTAGCTAGATTTTGCC75634GTMSVQEYAQKFNELARFCP			R	etr	otr	ans	pos	on	gag	pr	ote	in	(pf	am	acc	ess	ion	PF	037	32)	
IVRAQENFNWNAFKVAIKDRGATTTTTCCCTGAACATATTAGGAGGCAGAAATACAATGAGTTCACTAGATTTAACCAGGFFFPEHIRQKYNEFTRFNQGGAGGTACTATGTCTGTGCAAGAGTATGCCCAAAAGTTCAATGAGTTAGCTAGATTTTGCC75634GTMSVQEYAQKFNELARFCP	CTAI	TGT	TAG	AGC	TCA	GGA	AAA	СТЛ	TAA	TTG	GAA	TGC	TTT	TAP	GGT	TGC	TAT	ТАА	GGA	TA	75754
GATTTTTCCCTGAACATATTAGGAGGCAGAAATACAATGAGTTCACTAGATTTAACCAGG 75694 F F P E H I R R Q K Y N E F T R F N Q G GAGGTACTATGTCTGTGCAAGAGTATGCCCAAAAGTTCAATGAGTTAGCTAGATTTTGCC 75634 G T M S V Q E Y A Q K F N E L A R F C P	I	V	R	А	Q	Е	N	F	N	W	N	А	F	K	V	А	I	K	D	R	
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$																					
F       P       E       H       I       R       Q       K       Y       N       E       F       T       R       F       N       Q       G         GAGGTACTATGTCTGTGCAAGAGTATGCCCAAAAGTTCAATGAGTTAGCTAGATTTTGCC       75634         G       T       M       S       V       Q       E       Y       A       Q       K       F       N       E       L       A       R       F       C       P	GATI	TTT	CCC	TGA	ACA	TAT	TAG	GAG	GCA	GAA	ATA	CAA	TGA	GTI	CAC	TAG	ATT	ТАА	CCA	GG	75694
GAGGTACTATGTCTGTGCAAGAGTATGCCCAAAAGTTCAATGAGTTAGCTAGATTTTGCC 75634 G T M S V Q E Y A Q K F N E L A R F C P	F	F	Р	Е	Н	I	R	R	Q	K	Y	N	Е	F	т	R	F	N	Q	G	
GAGGTACTATGTCTGTGCAAGAGTATGCCCAAAAGTTCAATGAGTTAGCTAGATTTTGCC 75634 $G T M S V Q E Y A Q K F N E L A R F C P$																					
G T M S V Q E Y A Q K F N E L A R F C P	GAGG	STAC	TAT	GTC	TGT	'GCA	AGA	GTF	ATGC	CCA	AAA	GTI	CAA	TGA	GTT	AGC	TAG	ATT	TTG	CC	75634
	G	Т	М	S	v	Q	Е	Y	А	Q	K	F	N	Е	L	А	R	F	С	Р	

СТАА	TGT	TGT	GCC	AGA	TGA	GAG	SAGC	TAA	GGC	TCA	AAA	GTT	TGA	GGA	TGG	TTT	AGC	ATT	'TA	75574
N	V	V	Ρ	D	Ε	R	А	K	А	Q	K	F	Ε	D	G	$\mathbf{L}$	А	F	R	
GAAT	TCA	GAC	CAG	SACI	TGG	GGG	GAGC	AAC	TTC	TGC	AAC	TTT	TCA	GGA	AGC	TTA	TGC	TAA	GG	75514
I	Q	т	R	$\mathbf{L}$	G	G	А	т	S	А	т	F	Q	Е	А	Y	А	K	А	
CTTC	TAA	TAT	TGA	GAG	GAT	TTT	GAG	GCG	TGA	AGA	GGA	AGT	TAT	GGG	GAG	GAA	TAA	GAG	SAA	75454
s	N	I	Е	R	I	$\mathbf{L}$	R	R	Е	Е	Е	v	М	G	R	N	K	R	K	
AAGA	CCC	ACC	TAG	GCAA	CCA	AAA	TGA	CCA	TGG	AAA	TGA	CAA	GAA	ACC	TCG	АТА	TGG	GGG	та	75394
D	Р	Р	s	N	Q	N	D	Н	G	N	D	K	K	Р	R	Y	G	G	N	
					-															
ACAA	ТАА	ТАА	TGG	GGG	CAA	ТАА	TCA	CAC	таа	TGG	TGG	TGG	таа	ТТА	тса	AGG	GAA	TCG	та	75334
N	N	N	G	G	N	N	н	 Т	N	G	G	G	N	y	0	G	N	R	s	
		1,	U	U				-		U	U	U	1,	-	¥	U			U	
CCAA	ርመአ	CC 7	NCC	ምሮ እ	ccc	CAC	ነለመር	התתי	CCA	CCA	ACC	አመሮ			007	C 7 7	ccc	יייית כ	ጣጣ	75274
UCAA M	v		C	0	0000 C	DAD	c c	M			C	d d d	טשש. ס	m	.ссд о	N N		. I A C	.11 C	15214
IN	ĭ	Q	G	Q	G	ĸ	ъ	IN	Q	Q	G	5	ĸ	Т	Q	IN	Р	T	C	
0.000.0		0.000						1.00					200			<b>7</b> 7 <b>F</b>				75014
GTAG	JAAA	GTG	i'I'AA		AAG	iCCA		AGG	ATT T	TAC	CTG	TCA	AGG	AGA	.000	AAT		TTG	TT	/5214
R	K	С	Ν	K	S	Н	Р	G	F.	Т	С.	Q	G	ע -	Р	T	т	С	. Y	
										Z	ınc	kn	uck	le	(pi	am	acc	ess	101	n PF00098)
ATGC	TTG	TGG	AGA	GAA	AGG	GCA	TAA	GGC	TAA	TCA	GTG	TCC	CAA	GCG	TCA	GAA	TAA	TGG	SAC	75154
A	С	G	Ε	K	G	H	K	A	N	Q	С	Ρ	K	R	Q	N	Ν	G	Q	
AAAA	TGG	AAA	CAA	TGG	GGG	AAA	TAG	GAA	TGG	тса	TGG	GCC	TAA	TCA	GAA	CCA	GAA	TAA	CA	75094
N	G	N	N	G	G	N	R	N	G	Η	G	Ρ	N	Q	N	Q	N	N	Ν	
АТАА	CCG	TCC	СТА	CAA	CAA	CAA	CAA	СТС	TCA	AGG	TCA	AAC	TTC	GAA	TGC	TCA	AGG	GGG	GA	75034
N	R	Ρ	Y	N	N	N	N	S	Q	G	Q	т	S	N	А	Q	G	G	Ν	
ACAA	TAC	TCA	GCA	TAA	TGG	TCA	GAA	TAA	CAA	TCG	AGC	AAA	TGG	AGG	AAA	CAA	CAA	TCA	GA	74974
N	т	Q	Н	N	G	Q	N	N	N	R	А	N	G	G	N	N	N	Q	N	
ACGG	CAA	TGG	AAA	TGG	TGC	TCG	AGG	CAA	CAA	TGG	AAG	AAT	СТА	TGT	TAT	GAA	CCA	GAA	TG	74914
G	N	G	N	G	A	R	G	N	N	G	R	I	Y	v	М	N	0	N	Е	
-		-		-	 R	et r	ovi	ral	a.s	par	t.vl		- 0+0	ase	. (n	fam	- ac	ces		on PF08284
AAGO	יאכא	CAC	ממיזי		ית ממיזי	тст	יידעי			год Г Г Д П Д П	יד עכי יתיתיי	ירכיי דירריי	ירפיי	מממי	י רש מייר איי	ΤΔΔ		ידיבר	ידים. יידידי	74854
7000	AUA. ח			7000	M	17	1.01	m	-000 C		T	<u>т</u>	1001	M	<u> </u>	M		, I GC	v	19034
A	υ	Т	IN	А	IN	v	v	Т	G	Т	Г	Ц	v	IN	G	IN	Р	А	ľ	

АСТТ	GCT	TTT	TGA	TTC	TGG	GGC	GTC	TCA	TTC	TTT	CAT	AGC	TAG	TTC	ATT	TGT	TGA	AAA	GΤ	74794
L	$\mathbf{L}$	F	D	S	G	А	S	Н	S	F	Ι	А	S	S	F	V	Е	K	L	

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

TTCCTTGTAGTTCTCTATACCAAGACATACCCATTACCATATTAGGATCTGATTTGCCGG 74674 PCSSLYQDIPITILGSDLPA

CTGATCTTATTCAGTTTGACCTACCCGACTTTGATGTAATATTGGGAATGGATTGGCTTG 74614

CTAAGTATAGAGCTAGGATAGAGTGTCATACTCAAAAGGTGTCTCTTAGGGGGGCCAAAGG 74554 KYRARIECHTQKVSLRGPKG

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

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

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

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

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

CTTGGGGAGCACCAGTGTTATTTGTTCGAAAGAAGGATGGAACCTTGAGGTTGTGTATTG 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)

ATTACAGAGAGTTGAATAATGTCACAATAAGAATAAGTACCCATTGCCTAGGATTGAGG 74074

Ì	ATTT	ATT	TGA	TCA	АСТ	TAA	GGG	TGC	TGG.	ААТ	TTT	СТС	TAA	GAT	TGA	TTT	GAG	GTC	TGG	GΤ	74014
l	L	F	D	Q	L	Κ	G	А	G	Ι	F	S	K	Ι	D	L	R	S	G	Y	

Η	Q	$\mathbf{L}$	R	I	S	Е	Е	D	I	Р	Κ	Т	А	F	R	Т	R	Y	G
GGCA	ATTA	TGA	GTT	'CAC	AGT	GAT	GCC	ATT	TGG	АСТ	TAC	TAA	TGC	ACC	TGC	AGC	ATT	TAT	GG
Η	Y	Е	F	Т	V	М	Ρ	F	G	L	Т	N	А	Ρ	А	А	F	М	D
АТСІ	TAT	GAA	TAG	AAC	ATT	тса	GCC	GTA	TTT	AGA	TAG	ATT	TGT	GGT	GGT	GTT	CAT	AGA	ΤG
L	М	N	R	Т	F	Q	Ρ	Y	L	D	R	F	V	V	V	F	Ι	D	D
АТАТ	ATT	GGT	GTA	TTC	GAA	GGA	TAA	AGA	AGA	GCA	TGA	AGG	тса	TTT	AAG	GAA	AGT	TTT	GG
I	$\mathbf{L}$	V	Y	S	K	D	K	Е	Ε	Η	Е	G	Η	L	R	K	V	L	Ε
AGAT	ACT	TCG	AGA	GAA	AAG	GTT	GTA	TGC	TAA	GTT	ATC	AAA	ATG	TGA	GTT	TTG	GCT	TGA	GA
I	L	R	Ε	K	R	L	Y	A	K	L	S	K	С	Ε	F	W	L	Ε	K
AAGT	TGC	ATT	TTT	'AGG	TCA	TGT	GAT	TTC	GAA	GGA	AGG	TGT	TGC	TGT	AGA	TCC	ATC	AAA	GA
V	А	F	L	G	Η	V	Ι	S	K	Ε	G	V	A	V	D	Ρ	S	K	Ι
TACA	AGC	AGT	AAC	AGA	ATG	GGT	GAG	ACC	TAG	TAA	TGT	GAC	TGA	GAT	TAG	AAG	TTT	CTT	AG
Q	A	V	Т	Ε	W	V	R	Ρ	S	N	V	Т	Ε	Ι	R	S	F	L	G
							_												_
GACI	'TGC	TGG	СТА	CTA	TAG	GAG	GTT	TGT	GCA	AGA	TTT –	СТС	AAA	AGT	AGC	тса	ACC	TTT	GA _
Г	A	G	Y	Y	R	R	F	V	Q	D	F	s	K	V	A	Q	Р	L	т
																~			
CAAA	.T.T.T.	GAT	'GAA	IGAA	.AAC	AAC	TCG	ATT	TCA	GTG	GGA	'I'GA	GAG	GTG	'I'GA	GAA	AGC	TTT	TC
N	Г	М	K	K	т	т	R	F.	Q	W	D	E	R	С	E	K	A	F.	Q
												<b></b>				<b>—</b> — — — — — — — — — — — — — — — — — —			
AGGA	ATT	GAA	GCA	LAAG	ACT	TAC	TTC	AGC	ACC	AGT	TTT	GAC	A'I''I'	ACC	ATC	TGG	ATT	AGA	AG
E	Ц	ĸ	Q	ĸ	Ц	т	5	A	Р	v	Г	т	Ь	Р	5	G	Г	E	G
amma		~~ <b>~</b>	0.00	<b>m a</b> 7		000		<b>m a a</b>	~ ~ ~	maa	~~~~	200	<u>م</u> سرح	<b></b>	<b>۲</b> mm	<u> </u>	0.07	707	<b>m</b> 7
GTTI	TGA	.GGT	GTA	TAG C	TGA	CGC	TTC	TAA	GAA	rGG	GTT	AGG	ATG	TGT	ATT T	GAT	GCA	ACA	ATA
F.	Ę	V	Y	5	ט	А	5	ĸ	N	G	Ц	G	C	v	Ц	М	Q	Н	5
ر سە	COM		200	אחא	псс	mm.~	<b>~</b> ~ ~	707	<b>م</b> م	<b>m 7</b> 7	~~~	<b>~</b>	<b>п</b> <i>с</i> т	707	<i><b>C</b> 7 7</i>	<b>m</b> m *	000	<b>س</b> ه م	mа
GTAA	r. Teen	GGT 77	AGC	ATA	TGC	TTC	GAG	ACA	ACT	TAA	GCC	1.1.1A	TGA	ACA	GAA	1.1.4		TAC	TC
к	v	v	A	Y	A	S	к	0	1.	ĸ	Р	Ý	H:	0	N	Y	Р		н

ATGA	TTT	AGA	GTT	'AGC	TGC	TGT	AGT	ATT	CGC	ATT	GAA	AAT	TTG	GAG	GCA	TTA	TTT	GTA	TG	73234
D	L	Ε	L	A	A	v	V	F	A	L	K	I	W	R	Η	Y	L	Y	G	
GTGT	GTC	ATG	ТАА	GAT	TTT	CAC	TGA	TCA	TAA	AAG	тст	GAA	АТА	TAT	ATT	TAC	тса	GAA	GG	73174
V	S	С	K	I	F	Т	D	Η	K	S	L	K	Y	I	F	Т	Q	K	Е	
AGTT	GAA	CAT	GAG	ACA	GAG	GAG	ATG	GCT	TGA	АСТ	TAT	ТАА	GGA	TTA	TGA	TTT	AGA	GAT	тт	73114
L	Ν	М	R	Q	R	R	W	L	E	L	I	K	D	Y	D	L	Ε	I	L	
TGTA	TCA	TGA	GGG	ТАА	AGC	GAA	TAA	AGT	TGC	TGA	TGC	ATT	GAG	TAG	GAA	GAC	TAG	TCA	тт	73054
Y	Н	Ε	G	K	A	N	K	V	A	D	A	L	S	R	K	Т	S	Н	S	
CGAT	GAA	CAT	GAT	GGT	GTT	ATC	TGA	GAG	ATT	GTG	TGA	AGA	TTT	CAG	GAG	CAT	GAG	TTT	AG	72994
М	N	М	М	V	L	S	Е	R	L	С	Ε	D	F	R	S	М	S	L	Е	
<b>7</b> 7 0 0 0	<b>( ) (</b>		<b>a</b> a <b>a</b>	100		2.00	0.07	200	<b>m</b>	<b>7</b> mm	~ ~ ~ ~	паа	2.00	2 11 12	0.00	<b>00</b> 1	100	<b>a a</b>	<b>a</b> m	72024
AAG'I' V	M	GGA E	GCA O	AGG G	igca 0	AGT V	GGA E	AGC A	TCA 0	ATT L	GAA N	TGC. A	ACT L	ATG C	CGT V	GCA O	ACC P	CAC T	CT L	/2934
			~		2				~							2				
TATT	CGA	TGA	GAT	TCG	AGA	GAA	GCA	AAG	TAG	TGA	TGA	GTG	GAT	GGT	GAA	GAT	AAA	GAA	AA	72874
F	D	Ε	Ι	R	Ε	K	Q	S	S	D	Ε	W	М	V	K	Ι	K	K	М	
TGAA	AGA	AGA	TGG	AGT	TGT	CAT	CGA	GTT	TGA	CAT	TGA	TGA	AAA	TGG	TGT	TGT	GAA	GTA	CA	72814
K	Е	D	G	v	V	I	Е	F	D	I	D	Е	N	G	V	V	K	Y	к	
AGGG	AAG	ATG	GTG	TGT	TCC	TAA	.GGA	TGA	.GGA	GTT	AAA	AAG.	AAA	GAT -	TTT	GGA	AGA	AGC	тс	72754
G	R	W	C	V	Р	K	D	E	E	Г	K	R	K	T	Ь	E	E	A	н	
АТАА	TAC	TCC.	АТА	TTC	TGT	GCA	TCC	TGG	AGG	AGA	ТАА	АСТ	ТТА	ТАА	GGA	TTT	GAA	GCA	GC	72694
N	т	Ρ	Y	S	V	Н	Ρ	G	G	D	K	L	Y	K	D	L	K	Q	Н	
ATTT	TTG(	GTG	GAA		CAT M	'GAA	ACG	TGA	AGT	GGC	AGA	GTT	TGT	TGC	AAA	GTG	TTT	GAC	GT	72634
Г	vv	vv	К	IN	М	К	к	Б	v	А	Е	г	v	А	К	C	Ц	T	C	
GTCA	GAA	AGT	GAA	GAT	TCA	GCA	TAT	GAG	ACC	TGG	TGG	ААТ	GAT	GCA	ACC	TTT	AGA	AGT	GC	72574
Q	К	V	K	I	Q	Н	М	R	Ρ	G	G	М	М	Q	Ρ	L	Ε	v	Ρ	
0070	mma	~~~~	<u>م</u> س ح	007	000	<u>س</u> م	Int	egr	ase	CO	re	dom	ain	(p	fam	ac	ces	sio	n I	PF00665)
CGAG	T.T.G	GAA.	ATG	Aטטי ד	c TC	TAT T	e r	AAT M	GGA T	TTT F	TGT	GAT M	GGG C	ATT T	ACC P	ACT T	TAC	TAA	GT	12314
5	¥ ¥	1/	* *	Ľ	5	-	5	1.1	J	T,	v	1.1	J	ц	r		1	1/	5	

CAGC	TAA	GAA	TGC	CAT	'ATG	GGI	TAT	'AGT	'GGA	TCG	ATT	GAC	AAA	GTC	GGC	CAG	ATT	TAT	AG
A	K	N	A	I	W	V	I	V	D	R	$\mathbf{L}$	Т	K	S	А	R	F	I	A
CAAT	'GAA	GGA	TAC	ATG	GAG	TAT	GCA	ACA	GTT	GGC	TAG	TGC	ATA	TGT	GCG	AGA	GGT	TGT	ΤA
М	K	D	Т	W	S	М	Q	Q	L	A	S	A	Y	V	R	E	V	V	R
GACI	'GCA	TGG	AAT	'ACC	AAA	GGA	TAT	CGT	TTC	AGA	TAG	AGA	CTC	GAG	ATT	TTT	GTC	CAA	GT.
L	H	G	I	Ρ	K	D	I	V	S	D	R	D	S	R	F	L	S	K	F
ΓTTG	GGG	GAG	GTT	ACA	ACA	AGC	CTT	TGG	GAC	ATT	GCT	CAA	ATT	TAG	TAC	AGC	TTT	CCA	.CC
W	G	R	L	Q	Q	A	F	G	Т	L	L	K	F	S	Т	A	F	H	Ρ
CTGC	AAC	AGA	TGG	SACA	GAC	AGA	GAG	AAC	AAT	TCA	AAC	ATT	GGA	GGA	TAT	'GTT	'GAG	AGC	AT
A	Т	D	G	Q	Т	Е	R	Т	I	Q	Т	L	Е	D	М	Г	R	A	С
GTGI	'GAT	AGA	CTT	"TGG	GAGG	GATC	TTG	GGA	TGA	.TTA	TTT.	GCC	AAC	TAT	'AGA	GTT	TTC	GTA	TA.
V	Ι	D	F	G	G	S	W	D	D	Y	Г	Р	т	Ι	Е	F	S	Y	Ν
																			. –
ACAA	CAG	'T'T'A	TCA	CTC	CAAC:	CA'I		GAT	GGC	ACC	GTA	TGA	AGC	ATT T	'G'I'A	TGG	iGCG	AAA	AT
N	S	Y	н	S	S	T	K	М	A	Р	Y	E	А	Г	Y	G	R	K	С
2012		maa	mmm					<b>0</b> 7 m	N N C		<b>C D C</b>	<u>с</u> л п		mmm	1200			<b>7</b> 7 m	<b>C N</b>
JATE D	DAD'	TCC D	т. т	.GTG	M TTG	GAG C	TGA D	T	CAAG	TGA F	UGAC m	GAT	GAC	тт т	AGG	יטטטי	TGA	GAT	GA T
к	5	r	Ц	C	vv	5	D	T	5	Б	T	м	T	Ц	G	P	Б	м	Т
ኮጦር እ	767	770	יא ארי	1C 7 7		лст	יידיא כי	CCT	יידי איידי	יידירי א	CCA	CCA	Слт		ccc	ACC	יידירי א	767	ጥአ
LIGH F	LAGA F		.ддс т	V V		U DAI	DAL	т	т		RDD.	члы	м	DAD	שטט א	۸GC		лол	л. д
12	Ц	T	т	К	Q	v	К	Ц	Т	Q	Ц	11	м	К	А	А	Q	D	к
CACA		GGC	יידיים	CGC	מסמי	ጣሮ ል	GDD	ጥልር	AAG	GGA	СЪТ	GGA	<u>አ</u> ጥጥ	יידים	ССТ	יידים	CCA	GDD	GG
0	ĸ	Δ	v	Δ	люп. D	0	N	R	R	E	м	E	F	E	v	G	E	ĸ	Δ
¥	R		-		D	¥				-		-	-	-	·	U	-		
ንጥጥባ	יהכית		AGT	יርጥር	ימכר	יאאר	מממי	GGG	GGT	ירים	GAG	ልጥጥ	TGG	TAG	GAA	AGG		GTT	GA
т.	T.	ĸ	v	S	P	// ጠ10 ጥ	ĸ	.000 С	v	м	R	F	G	R	ĸ	G	ĸ	т.	s
Ц	Ц	к	v	D	1	1	к	U	v	11	к	1	U	к	к	U	к	Ц	U
ንግጥር	ארכ	ጥጥ እ	СЪл	ካጥርር	acc	'	TGA	GDT	ירייי	GGA	ACC	ልልጣ	יידים	GDD	ልርጣ	יאפר	ረጥጋ	ጣልር	ጃጥ
	P	v	т	C	יחכנ	v	F	т	т т	F	P	T	- G	v	10-11	٦٩٢	v	PUT	т
r	л	T	т	G	r	T	ല	T	Ц	ட	л	Т	G	л	v	А	T	к	ц
пасс	יריישיי	ACC	יעעי	ימכא	Curu	יממר	י עדיי	ጥርጣ	יררא	ጥልአ	CCT	Стт	יידירי	ጥርጣ	ርሞር	ጣር እ	ልሮጣ	ጥሮሮ	ממ
7571 7	т. Т	лсс п		. GGA	т С.	. AGC	. т.ст.ст М	17	U U	м м	1001	СТТ Б	ц	17	01C	- CA	T	тс <u>а</u>	הה ע
А	LL .	Р	IAT	E	Ц	А	IN	v	п	IN	v	г	п	v	5	Q	Ц	ĸ	ň

ААТА	TGT	CCA	rga:	FCC	FAC	CCAT	ГАТС	CAT	ГСА	ACC	TGA	AAC	CAT	<b>FGA</b>	ACT.	AGA	TGA	AAC	СТ	71674
Y	v	Н	D	Ρ	т	Н	I	I	Q	Ρ	Е	т	I	Е	L	D	Е	т	L	
TATC	CTT	rgao	GCA	ACG	CCC	AGTI	rago	GAT	гст	TGA	TAC	CAA	AAC	GAG	AAG	TAC	CCG	GAA	CA	71614
S	F	Е	Q	R	Р	V	R	I	L	D	Т	K	Т	R	S	Т	R	N	K	
			Ch	romo	o do	omai	in (	(pfa	am	acco	ess	ion	PF	003	85)					
AGGC	GGTA	AAA	АСТИ	AGTO	CAAG	GGTO	GTTA	ATGO	GTC.	AAG	TCA	AAC	FTC	<b>FGA</b>	AGA	GGC	TAC	TTG	GG	71554
Α	V	K	L	V	K	V	L	W	s	S	Q	Т	s	Е	Е	A	т	W	Е	
AAGC	CGA	AGA	rga:	ГАТС	GAA	AAA	CCGI	ATA	гсс	CGA	ACT	TTC	CCA	GCA	GGT.	ACG	CTT	GAG'	тт	71494
А	Е	D	D	М	K	N	R	Y	Р	Е	L	s	0	0	v	R	L	s	F	
L													-	-						
TCGG	GGA	CGA	AAC	гстт	гта	AGG	GGG	GTAG	GAA	TGT	GAT	АСТИ	AAC	rrr	ТТG	TTT	GTA	ТАТ	ТА	71434
G	D	Е	т	L	*															
-	-	-	-	-																
GTAG	CGA	GCG	4 TA	ACG	гта	AAG	гтсо	GAG	GAC	GAA	CTT	гст	רידי	AAG	GGA	GAG	TAG	ATG	ͲА	71374
АТАТ	CCC		 րդդո	TTA'	ГАА(	СТАТ	יתיתי	ГАТИ	4AA	АТТ	о ААТ'	 	 ዓጥጥ/	АТА	АТА	AGA	፦ ጥጥጥי	тат:	<u></u> АТ	71314
<u>атат</u>	<u>ата</u>	<u>א</u> רדיייייייייייייייייייייייייייייייייייי	TTG		<u>арт</u>	AAA(	37727	4 A A 7	ГСА	<u>таа</u>	ТААЛ	ATC	AGA	דידידי דידידי	тта	TGA	AAC	TGT	<u></u> ጥጥ	71254
<u>ттат</u>	GTA		<u>аат(</u>	TAG	<u>אידי א</u>		<u>ידידי</u> רידידי		AAG		<u>דידית</u>	CGA	<u>арт</u> а	CAA	<u>די ב ר</u>	<u>דיסי</u> ד דידידי	GAA		<u> </u>	71194
ATCA	GATT	 		<u>יידידי</u>	гт Gi	AAG				<u>איייי</u>	GGAZ	<u>סידיי</u>	rga'	<u>דידי</u>	TCA	GAT'	<u>ייייי</u> דידידי	TCG	TC	71134
CAAA	ACG		<u>אררי</u>		GAG					TGA	ם ביים ביים ביים ביים ביים ביים ביים בי	י <u>ד דר</u>	יידע	TTG		TGG	<u></u>	TGG	<u>10</u> A A	71074
AAGG	Δ		<u>т</u> тт	гтGr		Δ.Π.Δ.(	<u>ייייי</u>	<u>איריי</u>	<u>ר ד כ</u>	TAG	TGA		<u>ייי</u> יי		<u>דיסי</u>		<u></u>	<u>י מיד מ</u>	<u>π</u> Δ	71014
<u>тата</u>	<u>тас</u>	10/11	<u>יייי</u> ממר					<u>ייייי</u> אידיי	<u>ידד</u>	ACG	<u>таа</u> г		<u>ארטט</u>	דייייי <u>י</u>	<u>ттс</u>		<u></u>	TTG	<u></u> Ст	70954
AAGT	<u> </u>				<u>יסידר</u>		<u>אחת ה</u>		<u>יאר</u>	<u>רוביי</u> דוביי	TGTO	<u>።</u> ግጥጥי	TGA		<u>ד ד כ.</u> דידידי	CAA	CAA		<u> </u>	70894
	<u></u>	<u>יתיתי</u>		0101 0 0 01		<u>מתת</u>			2770	CCA			<u>רסיי</u>		<u></u>	<u>тст</u>			<u>ст</u>	70834
		<u>י ד ד י</u>	<u>ר ד ד ד</u> הששי	<u>ההה.</u> דריתי	<u>רככי</u>				<u>ידר</u>	<u>сал</u>		<u>דריי</u>		<u>דידכי</u>		<u>101</u>	<u>тет</u> Стра	<u>101</u>	<u>ст</u> Ст	70774
<u>тсст</u>	<u>עררי</u>	<u>רסיי</u>		<u>гсг.</u> гтаг		<u>י א</u> חי		<u>ייייי</u>					<u>יי</u> דער		<u>аст</u>				<u>21</u>	70714
	<u>тст</u>		<u>ר א ד</u>	<u>, 7 m</u> 7	<u></u>	<u>8 7 7 7</u>		NNN7	<u> </u>	777	7770	2777	<u>, , , , , , , , , , , , , , , , , , , </u>		<u>701</u>	ACA		CCN	<u></u>	70654
AAAG	<u>тсс</u> и			<u>יית גרו.</u>								<u>ייייי</u> ייייי	288			<u>пол</u>			<u>11</u> 77	70594
			יאמי					יד <u>ט זי</u>				<u>ייייינ</u> הששי			<u>אאא</u>				<u>77</u> 70	70534
	<u> </u>			7770											<u>сшу</u>				<u>m n</u>	70334
								<u></u>		<u>, , , , , , , , , , , , , , , , , , , </u>	<u>σι τ</u>							11G 170		70474
				<u>111</u>																70414
<u>TTTG</u>	TTG:								<u>, UU</u>			AGT.				CAN.		TGG		70354
		JDAG															TTG		<u>GT</u>	70294
GCTA	TGTO	JAT'A	A.L.G(		3.I	51'1' <i>I</i>	ATG'	rGA'		T'A'I'A	ATT.		2.TT.		T.G.T.	ATA'	7	T.G.T.		70234
CAGG	<u>AA'I'(</u>		ATA(	JA'I'A	<u>. T'T'</u>		ATGA	ACT/			ATT'17					ACT.		G'I'A		70174
ATTA	TCC(	TG(	GT7	A'I'A <i>l</i>	ATA'	<u>1'A'I'</u> /	ATG'	I'A'I'(	<u>A'I'A</u>	CTG	G'I'A'	1''1'G'	r'rG'	<u>1''1'A'</u>	TGA.	ATT	TGG	A'I''I''	<u>T'A</u>	70114
TAAT	GTA(		AT'		ACCI	ATCO	'I'A'I		AAA	'1''I'A'	rgC.		L'AT"	I'GA		TGT.	ACT'	1'AA'	TG	/0054
GGTA	TAA	CTAC	STG	rgt:	ľAA'	I'GA'	rGT <i>I</i>	AGCO	JAA	CGG	I'AT'	I'ATZ	AAG'	I'TG.	ACA	TAT.	ATT	GTA.	AC	69994
TCTA	TGA	GC	ГСТІ	AATO	<b>JAT</b>	GGAT	CATA	ATTO	ЗGТ	TAC	TAC	GAA'	<b>LL</b> L	rgg.	ACG	TTG.	ATT	AGT	AT	69934

${\tt TAAGTGGCTAAGTTGTGAAAAATATTATTTGAACTAAAGATGTTCCTGCTAATGTTAATG}$	69874
TGATGTTTGATGTGTCACAACTTTTAAAAATCTATTAATCACGTAAAGTGGAAATTGAGG	69814
GAATTATCCTGTGGATTTGGATCCTCCATAGGTGATGAACAGTACTTGATTTTATTATGA	69754
TACAACTCTTTATTGTCTTCCTCCTAATACTATTGGTGCGCATTGCGGATACCCATTAGA	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	ATGGCGTGGTACAGAAATTCAAGGTTTGTCTACAATGCTTTAAAACTCAACTTGCGTTCC	60
ORF19	ATGGCGTGGTACAGAAATTCAAGGTTTGTCTACAATGCTTTAAAACTCAACTTGCGTTCC	60
ORF20	ATGGCGTGGTACAGAAATTCAAGGTTTGTCTACAATGCTTTAAAACTCAACTTGCGTTCC	60
ORF21	ATGGCGTGGTACAGAAATTCAAGGTTTGTCTACAATGCTTTAAAACTCAACTTGCGTTCC	60
тк81-0	ATGGCGTGGTACAGAAATTCAAGGTTTGTCTACAATGCTTTAAAAACTCAACTTGCGTTCC	60

ORF18 AAAACATTTGGTACTATTCCAACTCCAAGAGTTCATTCGAATTCCTCATCTTTGTTTT	AC 120
ORF19 AAAACATTTGGTACTATTCCAACTCCAAGAGTTCATTCGAATTCCTCATCTTTGTTTT	AC 120
ORF20 AAAACATTTGGTACTATTCCAACTCCAAGAGTTCATTCGAATTCCTCATCTTTGTTTT	AC 120
ORF21 AAAACATTTGGTACTATTCCAACTCCAAGAGTTCATTCGAATTCCTCATCTTTGTTTT	AC 120
TK81-0 AAAACATTTGGTACTATTCCAACTCCAAGAGTTCATTCGAATTCCTCATCTTTGTTTT	AC 120

ORF18	AATCAATCTACTAATAAGTGTAGTGGGTTATTTGGGTCTGCAAAATCTGGGTATTTTAAT	180
ORF19	AATCAATCTACTAATAAGTGTAGTGGGTTATTTGGGTCTGCAAAATCTGGGTATTTTAAT	180
ORF20	AATCAATCTACTAATAAGTGTAGTGGGTTATTTGGGTCTGCAAAATCTGGGTATTTTAAT	180
ORF21	AATCAATCTACTAATAAGTGTAGTGGGTTATTTGGGTCTGCAAAATCTGGGTATTTTAAT	180
TK81-0	AATCAATCTACTAAGTGTAGTGGGTTATTTGGGTCTGCAAAATCTGGGTATTTTAAT	177

ORF18	GGGTTTAAACATCATCAAGAGATTAGCTCTTTCTCTGGTTTTGCAAGGAGAAATTATCAT	240
ORF19	GGGTTTAAACATCATCAAGAGATTAGCTCTTTCTCTGGTTTTGCAAGGAGAAATTATCAT	240
ORF20	GGGTTTAAACATCATCAAGAGATTAGCTCTTTCTCTGGTTTTGCAAGGAGAAATTATCAT	240
ORF21	GGGTTTAAACATCATCAAGAGATTAGCTCTTTCTCTGGTTTTGCAAGGAGAAATTATCAT	240
тк81-0	GGGTTTAAACATCATCAAGAGATTAGCTCTTTCTCTGGTTTTGCAAGGAGAAACTATCAT	237

ORF18	GGTGATAAAACCGAAGTAAGTGTTGAATCATGGCTGGAAAAATTCCTTGTTCCAATTGGA	300
ORF19	GGTGATAAAACCGAAGTAAGTGCTGAATCATTGCTGGAAAAATTACTTCTTCTTGCA	297
ORF20	GGTGATAAAACCGAAGTAAGTGTTGAATCATGGCTGGAAAAATTCCTTGTTCCAATTGGA	300
ORF21	GGTGATAAAACCGAAGTAAGTGTTGAATCATGGCTGGAAAAATTCCTTGTTCCAATTGGA	300
TK81-0	GGTGTTAAAAACCGAAGTAAGTGTTGAATTTCGGGTGGAAAAATTACTTCTTGGAATTGCA	297

ORF18	CTAATCTTGACTTTTGGTATACTTGGTTACCCTCATG <u>TGCACCCAGTAGTT</u>	351
ORF19	${\tt GTTGCAC}{-}{-}{\tt TAATCTTGA}{-}{-}{-}{{\tt TTGCTTACCGTCATG}}{\underline{{\tt TGCACCCAGTAGTT}}}$	342
ORF20	$\texttt{C} \texttt{TAATCTTGACTTTTGGTATACTTGGTTACCCTCATG} \underline{\texttt{TGCACCCAGTAGTT}}$	351
ORF21	$C TAATCTTGACTTTTGGTATACTTGGTTACCCTCATG\underline{TGCACCCAGTAGTT}$	351
тк81-0	$\texttt{C}{}{T}{A}{A}{T}{A}{T}{C}{T}{C}{G}{C}{A}{T}{T}{C}{T}{G}{C}{T}{T}{T}{C}{T}{T}{T}{T}{T}{T}{T}{T}{C}{C}{A}{C}{C}{C}{A}{G}{T}{A}{G}{T}{T}{T}{T}{T}{T}{T}{T}{T}{T}{T}{T}{T}$	351

ORF18	$\underline{\texttt{GTGCCA}} \texttt{TATACAGGAAGGAAGCATTATGTGCTTATGTCAACAACTCGTGAGAATGAAATT}$	411
ORF19	$\underline{\texttt{GTGCCA}} \texttt{TATACAGGAAGGAAGCATTATGTGCTTATGTCAACAACTCGTGAGAATGAAAAT}$	402
ORF20	$\underline{GTGCCA} TATACAGGAAGGAAGCATTATGTGCTTATGTCAACAACTCGTGAGAATGAAATT$	411
ORF21	$\underline{\texttt{GTGCCA}} \texttt{TATACAGGAAGGAAGCATTATGTGCTTATGTCAACAACTCGTGAGAATGAAATT}$	411
TK81-0	GTGCCATATACAGGAAGGAAGCATTATGTGATTTTGTCAACAACTCATGAGAATGAAAAT	411
	D-Fw	

ORF18	GGAGAAGTTGAGAAGCGGAAAATACAACCTGCTACACACCCTGATACTGATAGGGTTAGG	471
ORF19	GGAGAAGTTGAGAAGCGGAAAATACAACCTGCTACACACCCTGATACTGAGAGGGTTAGG	462
ORF20	GGAGAAGTTGAGAAGCGGAAAATACAACCTGCTACACACCCTGATACTGATAGGGTTAGG	471
ORF21	GGAGAAGTTGAGAAGCGGAAAATACAACCTGCTACACACCCTGATACTGATAGGGTTAGG	471
TK81-0	GGAGAATTTGAGAAGCGGAAAATACAACCTGCTACACACCCTGATACTGAGAGGGTTAGG	471

ORF18	TCAATATTCCAACACATTCTTGAATCACTGGAAAGAGAGATTAATCACCATGAACTCGAA	531
ORF19	TC <b>T</b> ATATTCCAACACATTATTGAATCACTGGAAAGAGAGATTAATCACCATGAACTCGAA	522
ORF20	TCAATATTCCAACACATTCTTGAATCACTGGAAAGAGAGATTAATCACCATGAACTCGAA	531
ORF21	TCAATATTCCAACACATTCTTGAATCACTGGAAAGAGAGATTAATCACCATGAACTCGAA	531
тк81-0	TCTATATTCCAACACATTCTTGAATCACTGGAAAGAGAGATTAATCACCATGAACTCGAA	531

ORF18	CTCGAACTCGAAAGAGATGAAACTTTCAAGGAGAAAACCATTTGGAAGGAGGAG	585
ORF19	CTCGAAAGAGATGAAACTTTCAAGGAGAAAACCATTTGGAAGGAGGAG	570
ORF20	CTCGAAAGAGATGAAACTTTCAAGGAGAAAACCATTTGGAAGGAGGAG	579
ORF21	CTCGAACTCGAAAGAGATGAAACTTTCAAGGAGAAAACCATTTGGAAGGAGGAG	585
тк81-0	CTCGAACTCGAACTCGAAAGAGATGAAACTTTCAAGGAGAAAACCATTTGGAAGGAGGAG	591

ORF18	ACAGTTGATGATAAAGATAGTAGGAAGAAGCATAGTGGGGGCTAAGATAACTACTAACCAT	645
ORF19	ACAGTTGATGATAAAGATAGTAGGAAGAAGCATAGTGGGGGCTAAGATAACTACTAACCAT	630
ORF20	ACAGTTGATGATAAAGATAGTAGGAAGAAGCATAGTGGGGGCTAAGATAACTACTAACCAT	639
ORF21	ACAGTTGATGATAAAGATAGTAGGAAGAAGCATAGTGGGGGCTAAGATAACTACTAACCAT	645
тк81-0	ACAGATCATGATAAAGATAGTAGGAAGAAGCATAGTGGGGGCTAAGATAACTACTAACCAT	651

ORF18	TTGGAAGGGATGAATTGGGAAATTTTCGTTGTTGATAAACCGTTGGTTG	705
ORF19	TTGGAAGGGTTGAATTGGGAAATTTTCGTTGTTGATAAACCGTTGGTTG	690
ORF20	TTGGAAGGGATGAATTGGGAAATTTTCGTTGTTGATAAACCGTTGGTTG	699
ORF21	TTGGAAGGGATGAATTGGGAAATTTTCGTTGTTGATAAACCGTTGGTTG	705
тк81-0	GAAGGGATGAATTGGGAAATTTTCGTTGTCGATAAACCGTGGGTTGAGTCCAGTTGT	708

ORF18	TTATTAGGTGGGAAGATTGTTGTTTACACCGGATTGCTCAACCATT-GCAACTCTGATG	763
ORF19	TTATTTGATGGGAAGATTGTTGTTTACACCGGATTGCTCAACCATTT-CAACTCTGATG	748
ORF20	TTATTAGGTGGGAAGATTGTTGTTTACACCGGATTGCTCAACCATT-GCAACTCTGATG	757
ORF21	TTATTAGGTGGGAAGATTGTTGTTTACACCGGATTGCTCAACCATT-GCAACTCTGATG	763
тк81-0	ATATTTGGTGGGAAGATTGTTGTTTACACTGGATTGCTCAACCATTTG-ATCTCTGATG	766

### Intron 1

ORF19	CTGAATT <u>GGCTACAATTATCGCGCATC</u> ACGTTGGGCATGCTGTGGCTCGACATGAGGCAG	808
ORF20	CTGAATT <u>GGCTACAATTATCGCGCATC</u> AGGTTGGGCATGCTGTGGCTCGACATGAGGCAG	817
ORF21	CTGAATT <u>GGCTACAATTATCGCGCATC</u> AGGTTGGGCATGCTGTGGCTCGACATGAGGCAG	823
тк81-0	CTGAATT <u>GGCTACAATTATCGCGCATC</u> AGGTTGGGCATGCTGTGGCTCGACATGAGGCAG	826
	Gre	

ORF18	AGGATTCGACAGCATTTTTCTGGTTGTTAATATCCCTCAACGTGATATTATTTAAAA	880
ORF19	AGCATTGGACAGCATTGTTCTGGTGGTCAATGTTAGGGTTCTACGTGACATTATTTGAAA	868
ORF20	AGGATTCGACAGCATTTTTCTGGTTGTTAATATCCCTCAACGTGATATTATTTAAAA	874
ORF21	AGGATTCGACAGCATTTTTCTGGTTGTTAATATCCCTCAACGTGATATTATTTAAAA	880
тк81-0	AGCATTGGACAACATTGTTGTGGTCGATACTGTTAGTGATATACATGACAATATTTCAAT	886

ORF18	${\tt TTCTATTTACTGAGCCTGAATCTGCCAATGCAAGATCAAAACTACT\underline{CTTAAGGCATCCTC}$	940
ORF19	${\tt TTCTATTTACTGCGCCTGAATTTGCCAATGCAAGATCAAAACTACT\underline{CTTAAGGCATCCTC}$	928
ORF20	${\tt TTCTATTTACTGAGCCTGAATCTGCCAATGCAAGATCAAAACTACT\underline{CTTAAGGCATCCTC}$	934
ORF21	${\tt TTCTATTTACTGAGCCTGAATCTGCCAATGCAAGATCAAAAACTACT\underline{CTTAAGGCATCCTC}$	940
тк81-0	${\tt atctatttactgcgcctgaatttgccaatgcaatatcaaaactact} \underline{{ctcaaggcatcctc}}$	946

	In	tron 2							
ORF18		<u>TCTTGC</u> AAAAA	GTTTGGAAGATTA	ATTCAGGCT	AGAGCTCC	АСААТТАСІ	GCCACGAAC	ΓA	1000
ORF19		TCTTGCAAAAG	GTTTGGAAGATTA	ATTCAGGCT	AGATTTCAT	ГСААТТАСІ	GCCACGAAC	ΓA	988
ORF20		TCTTGCAAAAA	GTTTGGAAGATT	ATTCAGGCT	AGAGCTCC	АСААТТАСІ	GCCACGAAC	ΓA	994
ORF21		TCTTGCAAAAA	GTTTGGAAGATT	ATTCAGGCT	AGAGCTCC	АСААТТАСІ	GCCACGAAC	ΓA	1000
тк81-0		TCTTGCAAAAG	GTTTGGAAGATT	ATTCAGGCT	AGATTTCAT	ГСААТТАСІ	GCCACGAAC	ΓA	1006
		D-Rv							

ORF18	${\tt TCTGCTTGTCCCTTGTTGGATTGTTTTCCTCGGTGTTTATTCTTTATTATGGTCGGA$	1057
ORF19	${\tt CCTTGCGATTGGGCTTTGTTGGATTGTCTTCCTTGGTGTGTTTATTCTTTATTTTGGTCGGA$	1048
ORF20	${\tt TCTGCTTGTCCCTTGTTGGATTGTTTTCCTCGGTGTTTATTCTTTATTATGGTCGGA$	1051
ORF21	${\tt TCTGCTTGTCCCTTGTTGGATTGTTTTCCTCGGTGTTTATTCTTTATTATGGTCGGA$	1057
TK81-0	${\tt CCTTGCACTTGGGCTTTCTTGGATTGTCTTCCTTGGTGTTTATTCTTTATTTTGGTCGGA$	1066

ORF18	${\tt AGGAAATAGAAGCAGATCACATTGGAGTGCTTCTGATGGCTTCTGCTGGATACGACCCGC}$	1117
ORF19	${\tt AGGAAATAGAAGCAGATCACATTGGAGTGCTTCTGATGGCTTCTGCTGGATACGACCCGC}$	1108
ORF20	${\tt AGGAAATAGAAGCAGATCACATTGGAGTGCTTCTGATGGCTTCTGCTGGATACGACCCGC}$	1111
ORF21	${\tt AGGAAATAGAAGCAGATCACATTGGAGTGCTTCTGATGGCTTCTGCTGGATACGACCCGC}$	1117
тк81-0	${\tt AGGAAATAGAAGCAGATCACATTGGAGTGCTTCTGATGGCTTCTGCTGGATACGACCCGC}$	1126

ORF18	GAGTTGCACCTCAAGTATATGACAAGCTTGCAAAGCCACTGGGCGACTGGAACTGTTTAG	1177
ORF19	${\tt Gagttgcacctcaagtatatgacaagcttgcaaagccactgggcgactggaactgtttag}$	1168
ORF20	${\tt Gagttgcacctcaagtatatgacaagcttgcaaagccactgggcgactggaactgtttag}$	1171
ORF21	${\tt Gagttgcacctcaagtatatgacaagcttgcaaagccactgggcgactggaactgtttag}$	1177
TK81-0	${\tt Gagttgcacctcaagtatatgacaagcttgcaaagccactgggcgactggaactgtttag}$	1186

ORF18	${\tt CAACTCATCCATTTGCAAGAATGAGAGGCAAAGTTGTTAGCTCGAGCTGATGTTATGAAGG}$	1237
ORF19	${\tt CAACTCATCCATTTGCAAGAATGAGAGCAAAGTTGTTAGCTCGAGCTGATGTTATGAAGG}$	1228
ORF20	${\tt CAACTCATCCATTTGCAAGAATGAGAGCAAAGTTGTTAGCTCGAGCTGATGTTATGAAGG}$	1231
ORF21	${\tt CAACTCATCCATTTGCAAGAATGAGAGCAAAGTTGTTAGCTCGAGCTGATGTTATGAAGG}$	1237
ТК81-О	${\tt CAACTCATCCATTTGCAAGAATGAGAGCAAAGTTGTTAGCTCGAGCTGATGTTATGAAGG}$	1246

ORF18	${\tt AAGCAGATAAGATATACAATGAAGTTGTAGCAGGACGTGCAATTCAAGGTCTTCAGTAA}$	1296
ORF19	AAGCAGATAAGATATACAATGAAGTTGTAGCAGGACGTGCAATTCAAGGTCTTCAGTAA	1287
ORF20	${\tt aagcagataagatatacaatgaagttgtagcaggacgtgcaattcaaggtcttcagtaa}$	1290
ORF21	${\tt aagcagataagatatacaatgaagttgtagcaggacgtgcaattcaaggtcttcagtaa}$	1296
TK81-0	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 bialaphos-resistance gene. A and B, anther morphology. C and D, images of Alexander's staining (scale bars; 20 μm).



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.

bvORF18/21 bvORF19 bvORF20 bvORF20L	MAWYRNSRFVYNALKLNLRSKTFGTIPTPRVHSNSSSLFYNQSTNKCSGLFGSAKSGYFN MAWYRNSRFVYNALKLNLRSKTFGTIPTPRVHSNSSSLFYNQSTNKCSGLFGSAKSGYFN MAWYRNSRFVYNALKLNLRSKTFGTIPTPRVHSNSSSLFYNQSTNKCSGLFGSAKSGYFN ************************************	60 60 60 59
bvORF18/21 bvORF19 bvORF20 bvORF20L	GFKHHQEISSFSGFARRNYHGDKTEVSVESWLEKFLVPIGLILTFG-ILGYPHVHPVVVP GFKHHQEISSFSGFARRNYHGDKTEVSAESLLEKLLLLAVALI-LIAYRHVHPVVVP GFKHHQEISSFSGFARRNYHGDKTEVSVESWLEKFLVPIGLILTFG-ILGYPHVHPVVVP GFKHHQEISSFSGFARRNYHGVKTEVSVEFRVEKLLLGIALIISHSGMIAFFYLHPVVVP *********************************	119 116 119 119
bvORF18/21 bvORF19 bvORF20 bvORF20L	YTGRKHYVLMSTTRENEIGEVEKRKIQPATHPDTDRVRSIFQHILESLEREINHHELELE YTGRKHYVLMSTTRENENGEVEKRKIQPATHPDTERVRSIFQHILESLEREINHHELELE YTGRKHYVLMSTTRENEIGEVEKRKIQPATHPDTDRVRSIFQHILESLEREINHHELELE YTGRKHYVILSTTHENENGEFEKRKIQPATHPDTERVRSIFQHILESLEREINHHELELE *******::***:*** **.***	179 176 179 179
bvORF18/21 bvORF19 bvORF20 bvORF20L	LERDETFKEKTIWKEETVDDKDSRKKHSGAKITTNHLEGMNWEIFVVDKPLVESSYLL RDETFKEKTIWKEETVDDKDSRKKHSGAKITTNHLEGLNWEIFVVDKPLVESSCLF RDETFKEKTIWKEETVDDKDSRKKHSGAKITTNHLEGMNWEIFVVDKPLVESSYLL LELERDETFKEKTIWKEETDHDKDSRKKHSGAKITTNH-EGMNWEIFVVDKPWVESSCIF ************************************	237 232 235 238
bvORF18/21 bvORF19 bvORF20 bvORF20L	GGKIVVYTGLLNHCNSDAELATIIA <u>HOVGH</u> AVARHEAEDSTAFFWL-LISLNVILFKILF DGKIVVYTGLLNHFNSDAELATIIA <u>HOVGH</u> AVARHEAEHWTALFWWSMLGFYVTLFEILF GGKIVVYTGLLNHCNSDAELATIIA <u>HOVGH</u> AVARHEAEDSTAFFWL-LISLNVILFKILF GGKIVVYTGLLNHCISDAELATIIA <u>HOVGH</u> AVARHEAEHWTTLLWSILLVIYMTIFQYLF .************************************	296 292 294 298
bvORF18/21 bvORF19 bvORF20 bvORF20L	TEPESANARSKLLLRHPLLQKVWKIIQARAPQLLPR-TICLSLVGLFSSVFILYYGRKEI TAPEFANARSKLLLRHPLLQKVWKIIQARFHQLLPRTTLRLGFVGLSSLVFILYFGRKEI TEPESANARSKLLLRHPLLQKVWKIIQARAPQLLPR-TICLSLVGLFSSVFILYYGRKEI TAPEFANAISKLLSRHPLLQKVWKIIQARFHQLLPRTTLHLGFLGLSSLVFILYFGRKEI * ** *** **** ***** *****************	355 352 353 358
bvORF18/21 bvORF19 bvORF20 bvORF20L	EADHIGVLLMASAGYDPRVAPQVYDKLAKPLGDWNCLATHPFARMRAKLLARADVMKEAD EADHIGVLLMASAGYDPRVAPQVYDKLAKPLGDWNCLATHPFARMRAKLLARADVMKEAD EADHIGVLLMASAGYDPRVAPQVYDKLAKPLGDWNCLATHPFARMRAKLLARADVMKEAD EADHIGVLLMASAGYDPRVAPQVYDKLAKPLGDWNCLATHPFARMRAKLLARADVMKEAD ******	415 412 413 418
bvORF18/21 bvORF19 bvORF20 bvORF20L	KIYNEVVAGRAIQGLQ 431 KIYNEVVAGRAIQGLQ 428 KIYNEVVAGRAIQGLQ 429 KIYNEVVAGRAIQGLQ 434	

FIGURE S9.– Comparison of amino acid sequences of five bvORF20-related genes between NK–198 and TK–81mm–O. Note that bvORF18 and bvORF21 are identical (see Fig. S5). Amino acid residues are numbered from the first methionine residue. Asterisks (\*) indicate positions that have a single, fully conserved residue; colons (:) indicate that one of the following 'strong' groups is fully conserved: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW. Points (.) indicate that one of the following 'weaker' groups is fully conserved: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, FVLIM, HFY (after CLUSTAL W package). HQVGH motifs in the bvORF20-related ORFs are underlined.

File S1.	List of gene	sequences	used fo	r phylogenetic	analysis.

Name of genes	Label in tree
AtPPR_1g01970 <sup>*1</sup>	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	GmREL 16
Glyma09g30620	GmRFL 18
Glyma09g30640	GmRFL 19
	GmRFL 20
Glyma09g30000	GmREL 21
Glyma09g30720	
Chuma00a20040	
Glyma09g30940	
Glyma09g39260	
	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
igilPoptr1 570963 eugene3.00130327	PtRFL7
igilPoptr1 571830 eugene3.00131194	PtRFL8
igilPoptr1 573736 eugene3.00190210	PtRFL 9
igi Poptr1 581077 eugene3.01250075	PtRFI 10
igi Poptr1 581161 eugene3 127570001	PtRFL11
igilPoptr1/594495/eugene3 00640083	PtRFL12
igilPoptr1/595453/eugene3.00700094	
	P+RFI 13
igilDoptr11505455100gopo2 00700006	PtRFL13
jgi Poptr1 595455 eugene3.00700096	PtRFL13 PtRFL14
jgi Poptr1 595455 eugene3.00700096 jgi Poptr1 595479 eugene3.00700120	PtRFL13 PtRFL14 PtRFL15
jg Poptr1 595455 eugene3.00700096 jg Poptr1 595479 eugene3.00700120 jg Poptr1 595494 eugene3.00700135	PtRFL13 PtRFL14 PtRFL15 PtRFL16
jgi Poptr1 595455 eugene3.00700096 jgi Poptr1 595479 eugene3.00700120 jgi Poptr1 595494 eugene3.00700135 jgi Poptr1 595495 eugene3.00700136	PtRFL13 PtRFL14 PtRFL15 PtRFL16 PtRFL17
jgi Poptr1 595455 eugene3.00700096 jgi Poptr1 595479 eugene3.00700120 jgi Poptr1 595494 eugene3.00700135 jgi Poptr1 595495 eugene3.00700136 jgi Poptr1 595506 eugene3.00700147	PtRFL13 PtRFL14 PtRFL15 PtRFL16 PtRFL17 PtRFL18
jgi Poptr1 595455 eugene3.00700096 jgi Poptr1 595479 eugene3.00700120 jgi Poptr1 595494 eugene3.00700135 jgi Poptr1 595495 eugene3.00700136 jgi Poptr1 595506 eugene3.00700147 jgi Poptr1 562855 eugene3.00070793	PtRFL13 PtRFL14 PtRFL15 PtRFL16 PtRFL17 PtRFL18 PtRFL19
jgi Poptr1 595455 eugene3.00700096 jgi Poptr1 595479 eugene3.00700120 jgi Poptr1 595494 eugene3.00700135 jgi Poptr1 595495 eugene3.00700136 jgi Poptr1 595506 eugene3.00700147 jgi Poptr1 562855 eugene3.00070793 jgi Poptr1 572581 eugene3.00140626	PtRFL13 PtRFL14 PtRFL15 PtRFL16 PtRFL17 PtRFL18 PtRFL19 PtRFL20
jgi Poptr1 595455 eugene3.00700096 jgi Poptr1 595479 eugene3.00700120 jgi Poptr1 595494 eugene3.00700135 jgi Poptr1 595495 eugene3.00700136 jgi Poptr1 595506 eugene3.00700147 jgi Poptr1 562855 eugene3.00070793 jgi Poptr1 572581 eugene3.00140626 AAM52339*4	PtRFL13 PtRFL14 PtRFL15 PtRFL16 PtRFL17 PtRFL18 PtRFL19 PtRFL20 PPR-592
jgi Poptr1 595455 eugene3.00700096 jgi Poptr1 595479 eugene3.00700120 jgi Poptr1 595494 eugene3.00700135 jgi Poptr1 595495 eugene3.00700136 jgi Poptr1 595506 eugene3.00700147 jgi Poptr1 562855 eugene3.00070793 jgi Poptr1 572581 eugene3.00140626 AAM52339*4 CAD61285	PtRFL13 PtRFL14 PtRFL15 PtRFL16 PtRFL17 PtRFL18 PtRFL19 PtRFL20 PPR-592 radish-Rf

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

<sup>\*2</sup>http://www.phytozome.net/search.php?show=text&org=Org\_Gmax\_v1.1

\*<sup>3</sup>http://www.phytozome.net/search.php?show=blast&method=Org\_Ptrichocarpa\_v2.2

\*4DDBJ/GenBank/EMBL dataase

\*5This 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); 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	MSWYRRTKLYFDSURRNINPKILPRSHVTSRINNPIGSSNPSAKFSSISS	50
Os_OMA1	MNYLKNSRSYLSRULR-HKPTGCPRLPPSPPLPQAPPAGYYFTSPS	45
BV_ORF19	MAWYRNSRFYYNAUKLNLRSKTFGTIPTPRVHSNSSSLFYNQST	44
Sc_OMA1	MKGTSGGFLKPVSF	25
At_OMA1	REVGLRSWTSLGRNTNRIAYNPFLSQPKRYYYVDRYQVRHFKPRGPGR	98
Os_OMA1	RPEAVRFGRVLLRSP-PPPPRPAQAPPSRYFYTSPQRQKVVHFNRRGSR	94
BV_ORF19	NKCSGLFGSAKSGYFNGFKHHQEISSFSGFARRNYHGDKTEVSV	88
Sc_OMA1	RVQLTRCYRYDNGPSYRRFNNGEYSQKSSFK	56
At_OMA1	WFQNPRTVFTVVLVGSVCLITLIVGNTETIPYTKRTHFILLSKPMPKLLG	148
Os_OMA1	WYHDPRKLTTVVVVSGGAAAAVYFGNLETVPYTNRTHLILSPPLPRQLG	144
BV_ORF19	ESWLEKFLVPIGLILTFCILGYPHVHPVVVPYTGRKHYVLMSTTRPNEIG	138
Sc_OMA1	SILLDKSSRKYLALLFGCCSLFYYTHLDKAEVSDRSRFIWVSRPLPLTIG	106
At_OMA1	ETQFEQIKKTYQGKILPATHPE <mark>SIRVR</mark> LIAKEVIDALQRGLSNERV	194
Os_OMA1	ESQFNNLKKELGPKILPPLHPDSIRVRLIASEVVRAVHRGLAGRHHDAFA	194
BV_ORF19	EVEKRKIQPATHPDTDRVRSIFQHILESLEREIN	172
Sc_OMA1	NYTYKSIWRQTQQE <mark>ILP</mark> PQHPL <mark>SI</mark> KIENIFMKIVE <mark>A</mark> AYKDPS	148
At_OMA1 Os_OMA1 BV_ORF19 Sc_OMA1	WSDLGYASTESSLGGG-SDKGVKEMEMAMSGEDTMTDMKWSKEDQVLD ADDASYGDISTDVVIKNHEAGAEDVMLGRSRGNKNASVAAAAQRDEEVLD RDETFKEKTIWKEETVD	241 244 198
At_OMA1	DQWIQKSRKKDSKAHAATSHLEGISWEVLVVNEBIVN-AFCLPAGKI	287
Os_OMA1	DRWVTESRDRGKARGAQPETRHLDGLNWEVIVVRDDLIN-AMCLPGGKI	292
BV_ORF19	DKDSRKKHSGAKITTNHLEGMNWEIFVVDKELVE-SSYLLGGKI	241
Sc_OMA1	VDNSLLDGIKWEIHVVNDETASPNAFVLPGGKV	181
At_OMA1	VVFTGLLNHFKSDAEVATVIGHEVGHAVARHVAEGITKNLWFAILQLV-L	336
Os_OMA1	VVFTGLLNHFKTDAEIATVLGHEVGHAIARHAAEMITKNLWFWILQIV-I	341
BV_ORF19	VVYTGLLNHCNSDAELATIIAHQVGHAVARHEAEDSTAFFWLLISLNVIL	291
Sc_OMA1	FIFSSILPICANDDGIATVLAHEFAHQLARHTAENLSKAPIYSLIGLV-L	230
At_OMA1 Os_OMA1 BV_ORF19 Sc_OMA1	-YQFV-MPDLVNTMSALFLRLPFSR -MQFIYMPDMINAMSTLLLKLPFSR	359 365 341 254
At_OMA1	KMEIEADYIGLLLASAGYDPRVAPTVYEKLGKLG	394
Os_OMA1	RMEIEADHIGLLVLGAAGYDPRVAPSVYEKLGKIA	400
BV_ORF19	FSSVFILYYGRK <mark>EIEADHIGVLLMASAGYDPRVAP2VYD</mark> KLAKPL	386
Sc_OMA1	QMETEADYIGLMIMSRACFQFQESIKVWERMANFEKQMNR	294
At_OMA1	GD-ALGDYLSTHPSGKKRSKLLAQANVMEEALMIYREVQAGRTGVEGFL	- 442
Os_OMA1	GDSTLSNYLSTHPSSKKRAQLLRQAKVMDEALRLYREVSSGQ-GTEGFL	- 448
BV_ORF19	GDWNCLATHPFARMRAKLLARADVMKEADKIYNEVVAGR-AIQGLQ	- 431
Sc OMA1	GGVVNMEFLSTHPASTRRIENMSKWLPKANEIYEQSDCSSMGNYYKSEFSM	4 345

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 Zn<sup>2+</sup> 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	MITAVIALISI GEWOKNELVNMV IFSOLHTFLIFSL
Candida_STE24	ST VA V LAGISTIGEWKLNGLPKMI TMMQGHLFLIFSL
Aspergillus_STE24	EV VAVISHELGHWSLGHTTKLFAIAQSHMFYIFAL
Coccidioides	EV VAV LSHELGHWSLSHTTKLFGIAQFHMFYIFAL
Neurospora	EV VAV LAHELGHWKLGHTTSLFGISQAHFFAIFSL
Os STE 24	EI VSVIAHELGHWKLNHTVYSFVAVQLLMFLQFGG
Hv ST E 24	EI VS VLA HELGHWK L NHTA Y SF VA V QL LT F MQ FG G
AT ST E 24	EI VAV IA HELGHWK L NHTT Y SF IA V QI LA F LQ FG G
Homo STE24	EV LAVIGHELGHWKLGHTVKNIIISQMNSFLCFFL
Mus STE 24	EV LAVIGHELGHWKLGHTVKNIIISOMNSFLCFFL
Bos STE 24	EV LAVIGHELGEWKLGETVKNI II SOMNSFLCFFL
Gallus STE24	EV LAVIGHELGHWKLGHTVKNTIISOMNSFLCFFL
	EV LAVIGHELCHWKLCHTVKNTVISOVNSFLCFFL
Tetraodon STE24	ET LAVIGUEL GUWKI. GUTVKNTVTSOMNSFLCFSL
Strongylogentrotug STE24	EVIAVIA HELCHWKI. CHNI. KNI. TI SOUNTI. LCIFI.
Drogophila	
OSCDMPL	ELOAVLARIELGELKCDRGVWLTFANILTMGAYS
Synechocystis	EIQAVMARIELGELKCERGVYLTLANIMVLAAGL
Crocosphaera	EIQGVMAHELGELKCEHGVYLTLANMMVLGASL
Anabaena	EIQAVIA HELGHLKCDHSVYLTPVNLLV LAASA
Thermosynechococcus	ELQAVIA HELGHLKCEHGVYLTIANLLL FAASQ
Synechococcus	EIQAVIA HELGHLKCNHGVYLTMANLLMLSTSL
Streptomyces	EMRAVIGHEVGHALSGHSVYRTILLFLTSLALRVA
Mycobacterium	EMRFVMG HELGHALSGHAVYRTMMHLLRLARSFG
Trichodesmium	EL KTVLAHELGHIKCGHPILNQMATWAMGIASAIT
htpX	EAEAVIA HEI SHIANGD MVTMTLIQGV VNTFVIFI
Haemophilus	EA EA VLA HE I SHISNGDMVTMALLQGVLNTFVIFL
Desulfitobacterium	EL EGVLAHEMAHIKNRD IL I STLAA - VMAGVI TTL
Nitrobacter	ELAGVIA HELAHIKHHD TLLMTITA - TIAGAI SML
Methanosarcina	EL E <mark>AVL</mark> AHEL SHVK N RD MAVLT IAS - FLSS VAFY I
Wolinella	EL AV VMG HE I AHAI A RHGA E RL SV S MA SE L GR NL I
Photobacterium	QLATVIGHEIGHVIA QHSNERLSRSQLANA GLELT
a-proteobacterium	QLASVMGHEIGHVIAEHGNERMSIATLSNLGLQIT
Nitrococcus	QLATVIGHEVGHVLAGHANERLSTNAATQTGLDLL
Pseudomonas	EI AAVMGHEI AHAL REHGREAM SKAYGVQVAS Q-I
Azotobacter	EI AAVMGHEI AHAL REHGREAL SKA YA VEMAK QGA
Chromobacterium	EL AAVIGHEISHALREHTRENM SOAYA OOMGLGLV
Bordetella	EL AAVIGHE I AHALREHARE RV SOOMATS I GL SVL
Flavobacterium	GLAMIT GUILAU ALANU GAORMIA OGOOIVGAA G
Ustilago	GLATVIGHEVAHOVARESAEKMSGYKVLLFGTFLL
Crvptococcus	GLATVIGHEIAHÖVARHPAERMSSMKVLFALGLLL
Omalp	GT AT V TA HE F AHOL A RH TA ENL SK A PT YS L LG LV L
Candida	GT ATVIS HEF AHOLA RUTA ENI, SKA PL YS LIG TIL
MR PR P - 1	OL SELUGUET ANAVI. GUAA EKA GMVHI. LD FLGMT F
Mil d	OL SELUCIET ALAVI. CHAAEKA SLVHI. DELCMIE
Ptrongylogentrotug	OL CTATA HEM ANALYAL NE SA EMA SEE FEED LEMINA
Gibberella	ALL ANY AGENT ASINT A SHARE FOL SA A WUCKLE TA COL
Aspergillus	CI. AN MACHET ANNUA DIATO POMON N PUTMON
y+WDI Vehetätttas	
	TAT AT A LOUIS AND A LOUIS A REAL AND A
	HIAT VIGHT VOHALA KHAA KMITKN LWFWI-LQIV
	OLATIIANU VIGHAVARNEAEDSTAFFWLLISLNVI
yrge	QLAS UMA HELSEVT Q KELA RAMED Q QR SA P LT WV G

File S4. See next page for the legend.

File S4. Multiple alignment of ~35 amino acid residues surrounding the Zn<sup>2+</sup> binding motif of peptidase M48 proteins, a protein family to which yeast OMA1 belongs. The position of the Zn<sup>2+</sup> 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; Coccidiodes, 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, Crocosphaera, Crocosphaera 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; aproteobacterium, 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.