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

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

*Laboratory of Genetic Engineering, Research Faculty of Agriculture, Hokkaido University, Sapporo, Hokkaido 060-8589, Japan, [†]Memuro Upland Farming Research Station, Hokkaido Agricultural Research Center, National Agriculture and Food Research Organization, Memuro, Hokkaido 082-0081, Japan, and [‡]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 and sequenced, and a solitary gene that was similar but not identical to *bvORF20* was found. Genetic features exhibited by 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 *Rf*s 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

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.
[1]Present address: Crop Breeding Research Division, NARO-HARC, Sapporo, Hokkaido 062-8555, Japan.
[2]Present address: Central Research Institute of Electric Power Industry, Abiko, Chiba 270-1194, Japan.
[3]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 *Rf*s seem to overcome the action of S factors in different ways, but the mechanisms are obscure. One group of *Rf*s regulates the expression of S-ORFs at the post-transcriptional 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 *Rf*s 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 Toriyama 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).

*Rf*s distinct from the PPR type, are known, but the current paucity of knowledge precludes further classification. Three non–PPR-type *Rf*s 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 copy-number variation between *Rf1* and *rf1*, suggesting that a common evolutionary mechanism(s) operates on plant *Rf*s.

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 $MgCl_2$.

Subcloning into a cosmid vector

Purified BAC-clone DNA was partially digested with *Sau*3A 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 $MgCl_2$, 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 *Xho*I and then partially filled to obtain a 5′-TC-3′ end (0.5 M Tris-HCl pH 7.5, 100 mM $MgCl_2$, 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 μg) 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 μg 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 ^{32}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 *Nco*I 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 *Sal*I and the other an *Nco*I 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 *Sal*I and *Nco*I 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 δ-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 *Bgl*II fragment containing *bvORF18* was obtained from cosmid clone 4F1 and subcloned into the *Bam*HI site of pBluescript. After verifying the nucleotide sequence, the fragment was excised as a *Pst*I–*Xba*I 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_4NO_3 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 2,3,5-triiodobenzonic 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 Mutator*-like 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

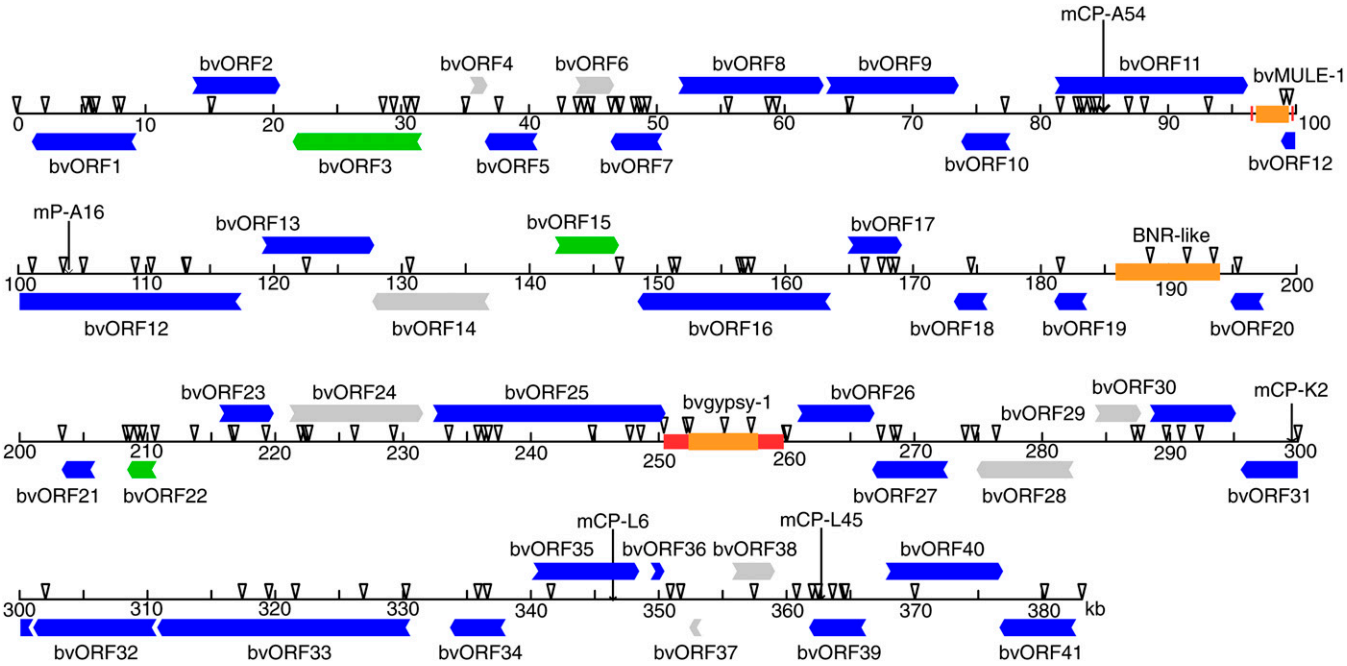


Figure 1 Organization of a 383-kbp chromosomal region of NK–198 deduced from two BAC clones and a cosmid clone. *Hin*dIII restriction sites are shown as triangles. Horizontal arrows indicate predicted genes and their orientation; intronic sequences are omitted. Gray arrows denote the absence of any homologous 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 *Rf*-like (*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 PPR-type *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			Transcripts[a]		
Name of ORFs	Locus name	Description[b]	*E*-value	Anthers	Leaves	Roots
bvORF1	At2g04940	Scramblase related	*e*-80	ND[c]	ND	ND
bvORF2	At4g33260	Putative cdc20 protein	0	ND	ND	ND
bvORF3	At5g17210	Unknown function	5*e*-44	ND	ND	ND
bvORF4	NA[d]	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)	4*e*-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	2*e*-63	ND	ND	ND
bvORF11	At2g34780	MEE22, EMB1611, etc.	4*e*-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	+[e]	+	+
bvORF14	NA	No hit	NA	+	+	+
bvORF15	At3g03150	Unknown function	3*e*-15	+	+	+
bvORF16	At5g42310	Pentatricopeptide repeat (PPR-like) superfamily protein	4*e*-94	+	+	+
bvORF17	At3g49010	60S ribosomal protein L13	8*e*-82	+	+	+
bvORF18	At5g51740	Peptidase M48 family protein	6*e*-62	+	+	+
bvORF19	At5g51740	Peptidase M48 family protein	4*E*-52	+	+	+
bvORF20	At5g51740	Peptidase M48 family protein	8*E*-61	+	+	+
bvORF21	At5g51740	Peptidase M48 family protein	6*E*-62	+	+	+
bvORF22	At3g50170	Unknown function	2*E*-71	+	−[f]	+
bvORF23	At5g48620	Disease resistance protein (CC-NBS-LRR[g] class) family	*e*-107	+	+	+
bvORF24	At5g51740	Peptidase M48 family protein	8e-06	+	+	+
bvORF25	At5g35450	Disease resistance protein (CC-NBS-LRR class) family	*e*-100	+	+	+
bvORF26	At1g58390	Disease resistance protein (CC-NBS-LRR class) family	*e*-107	+	+	+
bvORF27	At2g04620	Cation efflux family protein	*e*-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	2*e*-31	+	+	+
bvORF32	At4g27870	Vacuolar iron transporter (VIT) family protein	7*e*-35	+	+	+
bvORF33	At3g02580	Brassinosteroid biosynthetic enzyme		+	+	+
bvORF34	At5g24680	Peptidase C78, ubiquitin fold modifier-specific peptidase 1/2	2*e*-39	+	+	–
bvORF35	At3g49590	Autophagy-related protein 13	4*e*-96	+	+	+
bvORF36	At5g24660	RESPONSE TO LOW SULFUR 2 (LSU2)	2*e*-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	5*e*-65	ND	ND	ND
bvORF40	At5g24630	BRASSINOSTEROID-INSENSITIVE4 (a protein that forms part of the topoisomerase VI complex)	3*e*-36	ND	ND	ND
bvORF41	At5g24620	Pathogenesis-related thaumatin superfamily protein	2*e*-76	ND	ND	ND

[a] Summary of Figure S1.
[b] Descriptions from TAIR (http://www.arabidopsis.org/).
[c] No data.
[d] Not applicable.
[e] Detected.
[f] Not detected.
[g] 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

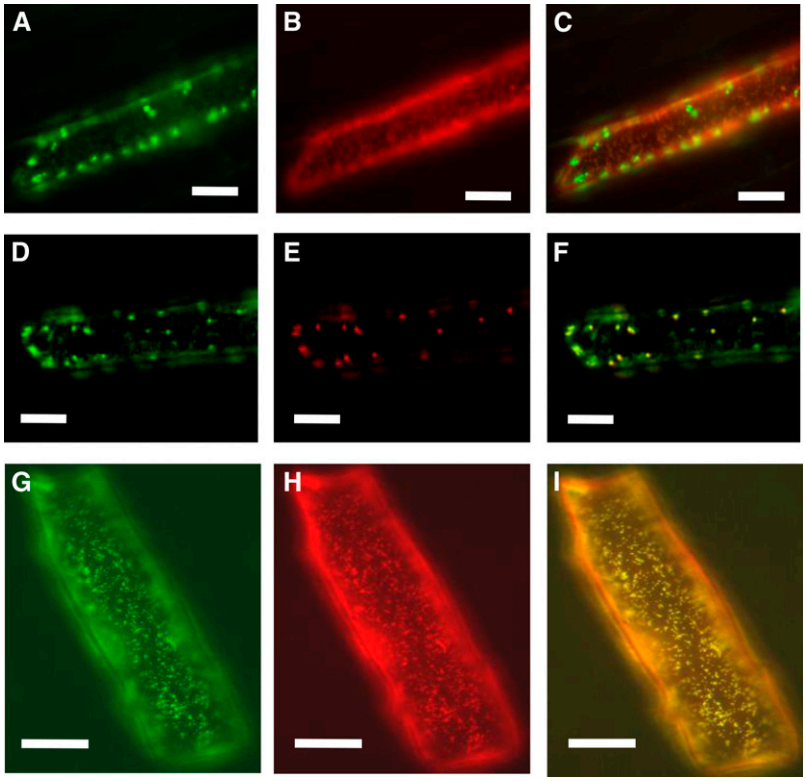


Figure 2 Images of fluorescent signals obtained from transient expression tests. (A–F) Images of epidermal cells of Welsh onion sheath. (G and H) Images of epidermal cells of onion bulb scales. Bars, 50 μm. A and D are green fluorescence images of bvORF16-GFP; B and H are red fluorescence images of mitochondria-targeted 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 *Hin*dIII 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 *Hin*dIII 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

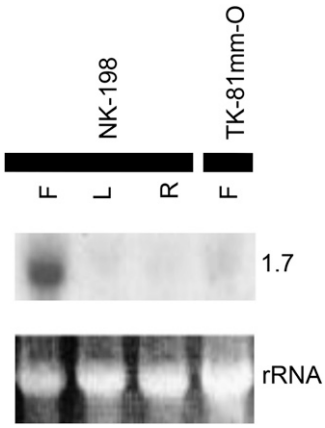


Figure 3 RNA gel blot analysis of the 3′-UTR of *bvORF19* hybridized with total RNAs from flower buds (F), leaves (L), and roots (R) of NK–198, and from flower buds of TK–81mm–O. Sizes of signal bands are indicated in kilobases. Images in the bottom row show ethidium-bromide (EtBr-) rRNA after gel electrophoresis.

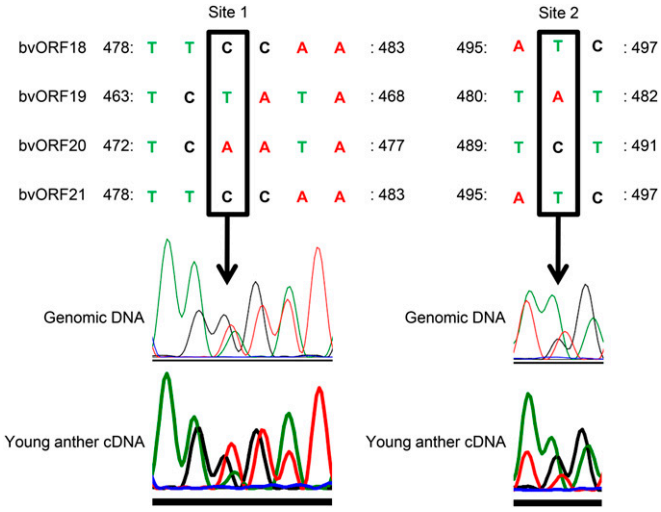


Figure 4 Polymorphic sites in the PCR targets of the quadruplicated genes in NK–198 and electrophoregrams obtained by direct sequencing. The original electrophoregrams were converted to complementary images on the sequencing platform (ABI3130). Red, green, and black lines indicate the signal peaks of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

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_1 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 fertility (Figure 5G). This partial-fertile phenotype was indistinguishable from that of the F_1 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_1 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*,

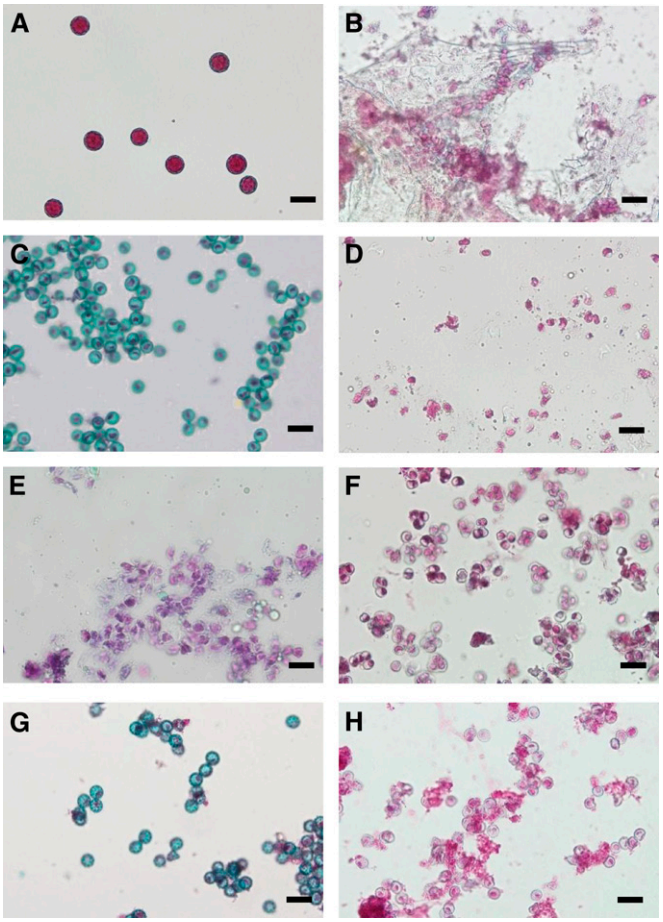


Figure 5 Photographs of anther contents from transgenic and control sugar beets. A–H are images of Alexander's staining. Bars, 20 μm. (A) Anther contents of a maintainer line, NK–219mm–O. (B) Anther contents of a CMS line, NK–219mm–CMS. (C) Anther contents of an F_1 plant (NK–219mm-CMS × NK–198). (D) 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.

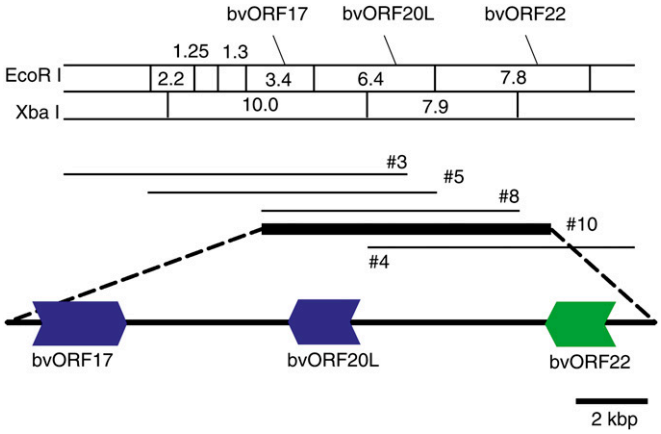


Figure 6 Physical map of the chromosomal region containing *bvORF17*, *bvORF20L*, and *bvORF22* of TK–81mm–O. Sizes of restriction fragments are shown in kilobase pairs. Five recombinant phage clones are indicated. Gene organization deduced from the nucleotide sequence of clone no. 10 is indicated below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

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 *Eco*RI and *Xba*I enabled us to assemble these clones into a contig of ~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 (hereafter 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

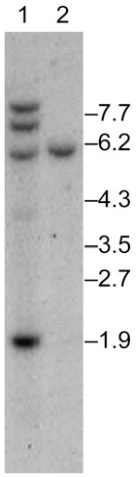


Figure 7 DNA gel blot analysis of the 3′-UTR of *bvORF19* hybridized with total cellular DNA from NK–198 (lane 1) and TK–81mm–O (lane 2). *Hin*dIII restriction endonuclease was used. Size markers are shown on the right (in kilobase pairs).

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_1 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 protease (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 *Rf*s. 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 mitochondria-targeting PPR protein. Theor. Appl. Genet. 108: 1449–1457.

Alexander, M. P., 1969 Differential staining of aborted and non-aborted 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 *RETROGRADE-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 male-sterile 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 glycine-rich 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 glycine-rich 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 repeat-containing 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 $tRNA^{Cys}$ 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 trichocarpa* (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 Sequence-specific 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

DOI: 10.1534/genetics.112.145409

TABLE S1 Primers used in this study

Purposes	Name of primers	Nucleotide sequences
cDNA sequencing of bvORF18/19/20/21	D-Fw	5'-TGCACCCAGTAGTTGTGCCA-3'
	D-Rv	5'-GCAAGAGAGGATGCCTTAAG-3'
	Gre	5'-GATGCGCGATAATTGTAGCC-3'
Generating hybridization probe of bvORF19 3' UTR	3'-FW	5'-AGCTTGCAAAGCCACTGGGCGA-3'
	3'-RV	5'-GGAACCAAATTAGATTGAATTAACAAGTGG-3'
bvORF16-GFP construction	GFP-ORF16-FW	5'-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’

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

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

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

```
CACTTTTGAGCAACTCACAATTTTATATATACATTACAAGTAATTAAAAATAAAGTATTA 100440
TATGGGTAACTTATACATTGATAAGGGTAACTAATAAGATAATTTGAGCAACTAATGTTT 100380
TTATATGTACATAACAAGTAAATAATAACACACGTTACATTGACTTGATTACACAAAGGA 100320
TAACTTATACATTGATAAAGACAACTAATAAGATAAGTTGAGCAACTAAAGATTTTACAT 100260
TGACTTGGTTAGATAAAGGATAACTTATAGATAGATAAAGGTAACTAATAAGATAATTTG 100200
AGCAACTAACAATTTTATATATACATTACAAGTAATTAAAATACAAAAAGTAAAGTATTA 100140
TATGGATAATTAAAAAAAAAAAGTAAAATATTATATGAGGAACTATGCAAATTCCTACTA 100080
ATGTACATTGACTTGGTTCAGTTAAACTTGTTGAAACATGTAATCTTGTAATCCTAGTGG 100020
AGTAATCTATTTACTTTTAGATTAACTTAAGCATAGTTTGGGATGGCCTATATGATGTTA 99960
ATAAGACATCACAACTAATCTCCAACGTTGTTTATATTAATGTGCATGTCATATGAAAAT 99900
GTCATAGACAAGTTCCAATAAGTCTAGATATGTCAAATTTCAGTTTACTAAATTATTATT 99840
GATTTATGTTCTTGTTGTTCCAGTAATCTACAATAGAATGGAAGATGAAGAATTTGCTCC 99780
AACTCCTTGCATTGATGCAACTCCAACTCCTTGTGTTGATACAACTCCAACTAATGCTAC 99720
TCAAACTTCTAATGCTCAAACAACAAGATCCACATTCACTCCTCGTCCATGCTACACACC 99660
TAGAGGTTCAAAAGAATGGATCCCTTGCTGCCCTCCTGAGTTAAAACCTACTGTGGGTAT 99600
 R  G  S  K  E  W  I  P  C  C  P  P  E  L  K  P  T  V  G  M

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

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

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

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

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
TGCTCACAAGAAGTTCATATTTGATAATTCAAGATTGAATGTTGGACCAAACAAATCTTT 99180
 A  H  K  K  F  I  F  D  N  S  R  L  N  V  G  P  N  K  S  F

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

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

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

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

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

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

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

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

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

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

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

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

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

TTTCAGAGTCTGAGAACAACTTTTTCACTTTGTTTACAAATGCAAATCTTCTTCTAGTTG 98340
  S  E  S  E  N  N  F  F  T  L  F  T  N  A  N  L  L  L  V  E
```

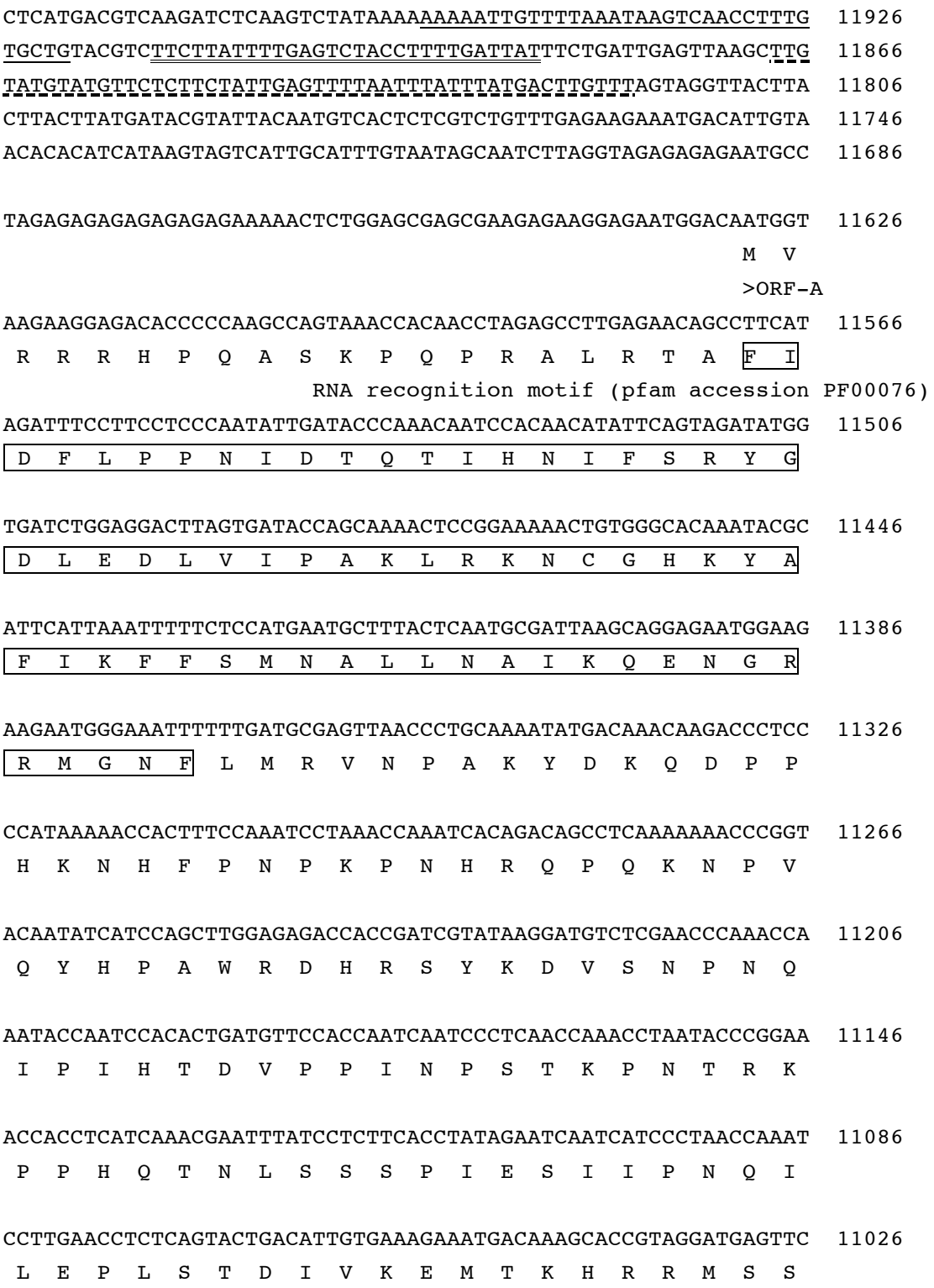

H. Matsuhira et al., FIGURE S2

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

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

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

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

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

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

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

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

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

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

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

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

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


H. Matsuhira et al., FIGURE S2

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

ATAAAAAGAAAAACCTAGAATGAGGTAAGTTTAGAATAATATATTATTTATACATCACTA 97440
  K   K   K   N   L   E   *

TCAAATGAGAGTTTCTTTTCATTCATTTATTAATAAATTATTTATCACAATTTTTGTTTC 97380
TTTTGCAGTTACATGGAGGATTGTGATACATGGTGATGACAAGAATCATGATGATCAAAA 97320
GGGGAGCAAGCAGATTTTACATCACACTTTTTGGGGAATTTTTGTTTGGTTTTATAAAGT 97260
GTTACCTACAAAAAATAAAGTGTATGTGTTCAAAGGAGCTGGCAAACATTTAAAAATTAT 97200
GGATTATTCTTTTGAAACGATTTCACTTGAAGGATTTTGATGGCTTAAGCACACTTTTGC 97140
AGTCTATCTCAAAAATTGTAGGTTTTGTGAAACTATATTCCATGGACACTGCAATTGATA 97080
TTGTAATTGGAAAGAACTTGATTATTCTGTAAAGTTTGCAAGTTATAACTTTTTCTTCTT 97020
AGGTTGGAAAGAACTAGATTTTGTAACTTTTTTTTCTTTTACTATTTGAGGTGACTGGAA 96960
TCCACCTTTTTGTGTTACCTTTTAGTGCATTCTTTGACTAGGACACTGATTCATTGGTAA 96900
TTATTATGTAAAAACCCCTATAAACAGTTACTAAAAAGAAGTTATAGTCAACCATACATT 96840
TACGAGAGAGTATTAATCTTTTTAGTTTGATTTTATGTATTATTTTCAAATATCTTTTTA 96780
TATGTACATAACAAGTAAATAATTTCACACGTTACATTGACTTGATTACACAAAGGATAA 96720
CTTATACATTGATAAGGGTAACTAATAAGATAAGTTGAGCAACTAAAGATTTTTACATTG 96660
ACTTGGTTAGCCAAAGGATAACTTATAGATAGATAAGGGTAAATAATATGATAATTTGAG 96600
CAACTAACAATTTTATATATACATTACAAGTAATTAAAATACAAGAAGTAAAGTATTATA 96540
TGGATAACTTATAGATAGATAAGGGTAACTAATAAGATAATTTGAGCAACTAACAATTTT 96480
ATATATACATTACAAGTAATTAAAATACAAGAAGTAAAGTATTATATGGATAACTTATAC 96420
ATTGATAAGGGTAACTAATAAGATAATTTGAGCAACTAACGATTTTATATGTACATAACA 96360
AGTAAATAATAACACACGTTACATTGACTATCTCCTTTCATCTTTATCAAATAAAAAATA 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.



H. Matsuhira et al., FIGURE S3

```
TAGGGTCCTTGGGGAAGACACGGAGAGAATAAGGGACCAAGTGGAACTTGTGGAACTAGA  10966
 R  V  L  G  E  D  T  E  R  I  R  D  Q  V  E  L  V  E  L  E

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

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

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

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

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

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

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

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

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

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

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

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


H. Matsuhira et al., FIGURE S3

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

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

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

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

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

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

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

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

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

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

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

CAAAAACCGCCTCTCCTCCTCCTCCTCCAATCCCCCACCTCTTCCTCCCTCTCCCTCCCC  9526
 K  N  R  L  S  S  S  S  S  N  P  P  P  L  P  P  S  P  S  P

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


H. Matsuhira et al., FIGURE S3

```
AGGAATGAAATTCAATGGTGAACTATCAGATCTACAAGACAAAATTGTTGGAATTTTGTC  9406
 G  M  K  F  N  G  E  L  S  D  L  Q  D  K  I  V  G  I  L  S

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

TTCCATGTTAATCTCGTGGAATGTCAGGGGCCTCGGAGCATGGCCTAAAAGAAATGTTCT  9286
CAAAAAGTTACTACTCCTTCATGACCCCATGATAGTATTCATCCAAGAATCCAAACTGGA  9226
ATGTATTCCTTCTAAATTGCAAAAATCAATTTGGTGTGATGATGACCTCAGCCTCTGTAT  9166
CAGTCCATCAAACGGATCCTCTGGAGGATTAATCTCCCTATGGAGACCCTCAAAATTTCA  9106
TCTGGTTTCCAGTAGAATCGAATCACAATGGATCGCAATGGAAGGAATGGTGGTGAGGGA  9046
                                      M  E  G  M  V  V  R  E
                                      >ORF-B
AAATTTTCAATGCCTTCTCATAAATATTTATAACTCCTGTGATGCTTCGACTAGATCAGA  8986
 N  F  Q  C  L  L  I  N  I  Y  N  S  C  D  A  S  T [R  S  D]
Endonuclease/exonuclease/phosphatase family (pfam accession PF03372)
CACATGGAACCATATAGAGGATTTTTGCAGAAACTCACACTTACCTCTTCTAATAGCGGG  8926
[T  W  N  H  I  E  D  F  C  R  N  S  H  L  P  L  L  I  A  G]

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

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

TGGATACTTTACATGGTTTAGGGGTCAATCAAAATCAAAGCTGGATAGAATTCTTGTCCA  8746
[G  Y  F  T  W  F  R  G  Q  S  K  S  K  L  D  R  I  L  V  Q]

ACCAGATTGGATTCTGAAATTCTCATTCCTTAATGCCTCCATCCTCAAAAGGAGTATCTC  8686
[P  D  W  I  L  K  F  S  F  L  N  A  S  I  L  K  R  S  I  S]

GGATCATTGCCCTCTTGTACTGAAGTCGCAATCTAAGGACCGGGGACCGAAACCTTTCAG  8626
[D  H  C  P  L  V  L] K  S  Q  S  K  D  R  G  P  K  P  F  R

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

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

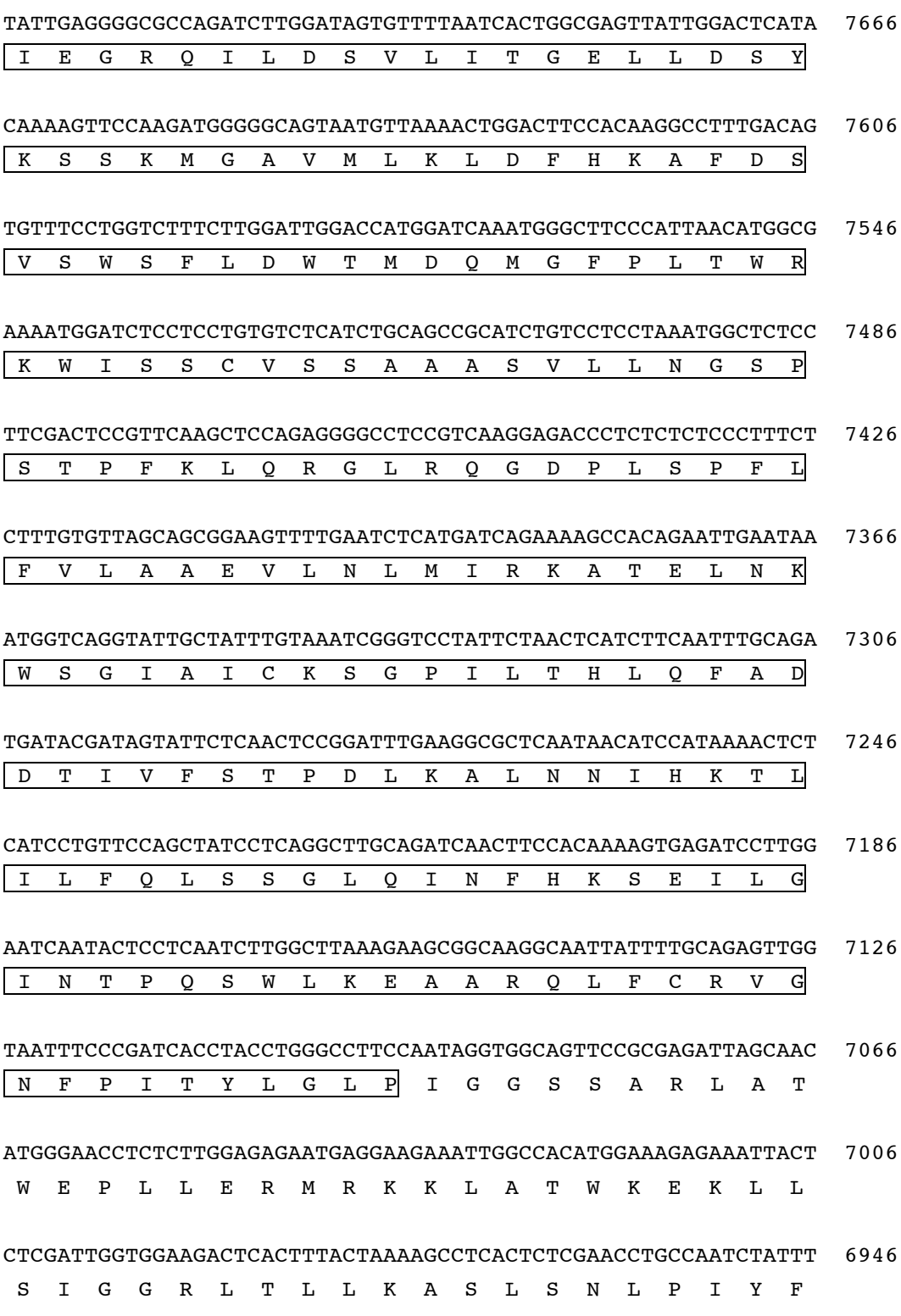

H. Matsuhira et al., FIGURE S3

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

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

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

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

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

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

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

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

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

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

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

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

RNA-dependent DNA polymerase (pfam accession PF00078)

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


H. Matsuhira et al., FIGURE S3



H. Matsuhira et al., FIGURE S3

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

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

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

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

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

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

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

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

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

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

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

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

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


H. Matsuhira et al., FIGURE S3

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

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

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

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

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

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

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

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

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

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

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

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

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


H. Matsuhira et al., FIGURE S3

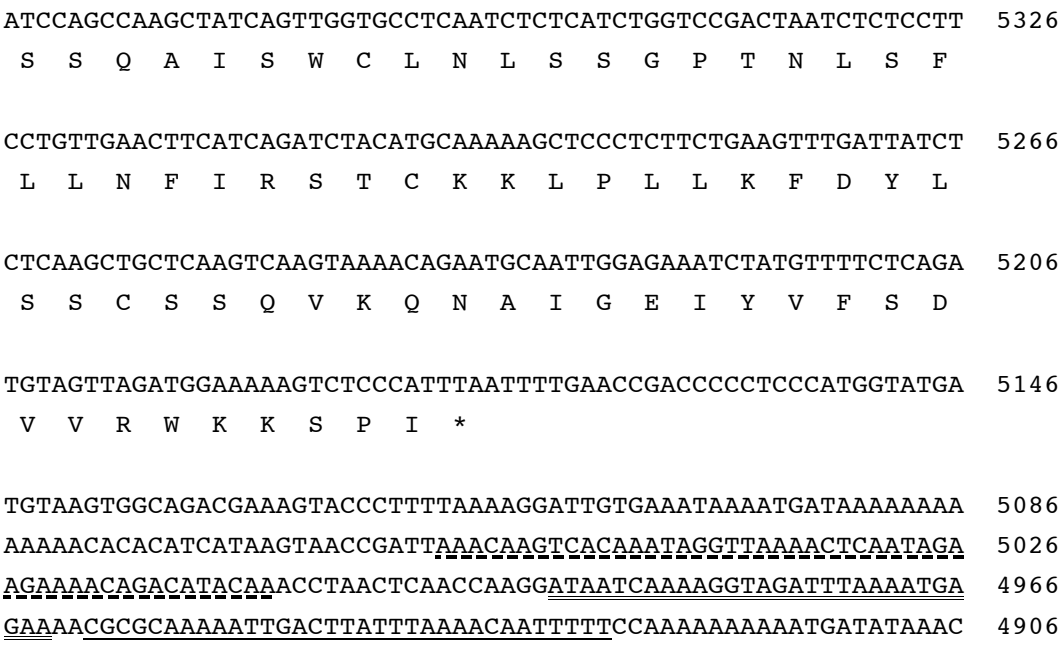
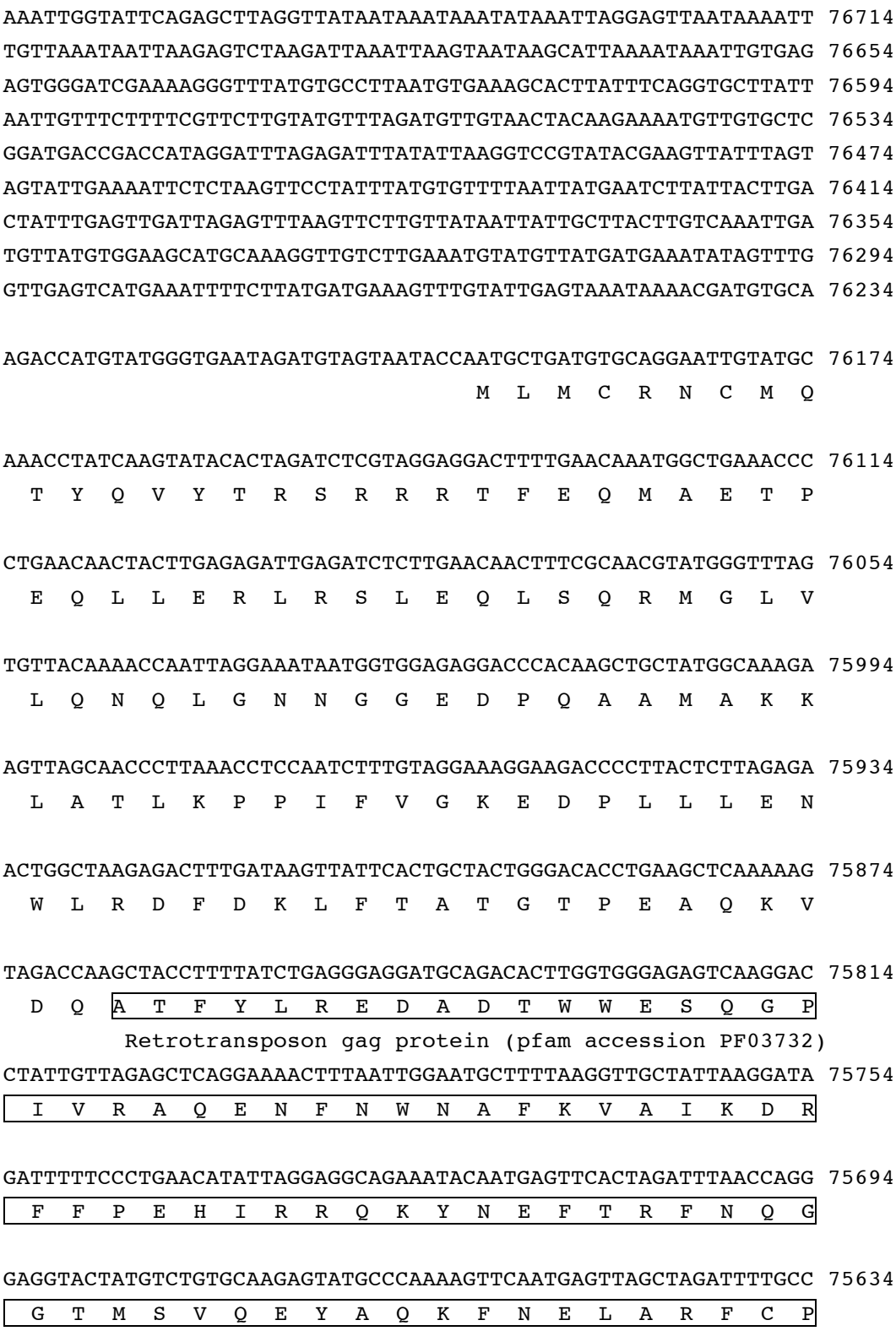


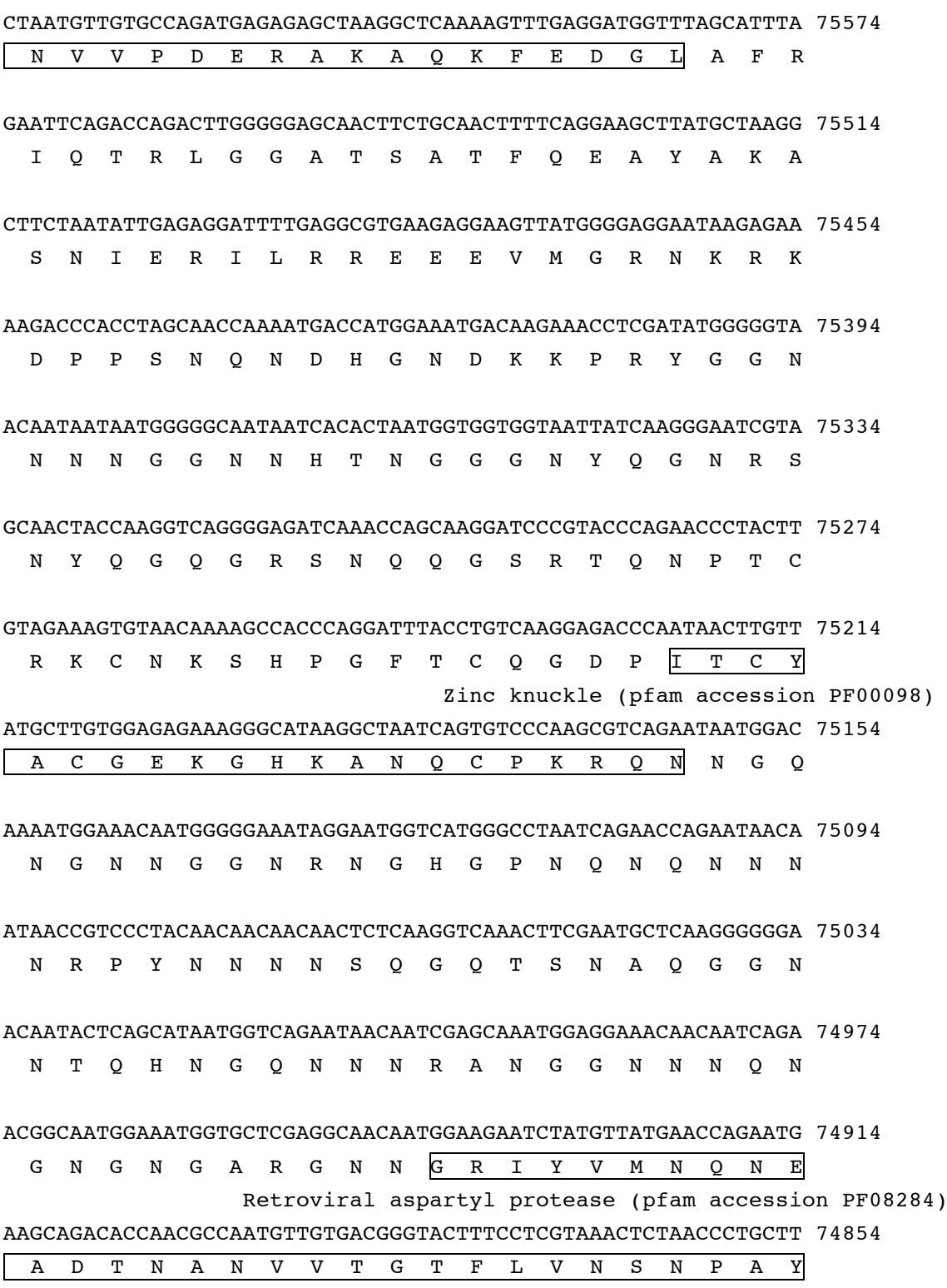
FIGURE S3.– Nucleotide sequence of the *BNR-like* element identified in this study. Numbers of nucleotide residues are coordinated with those of BAC clone 33E19. Each pair of palindromic sequences is indicated by single, double, or dotted lines. Putative domains are boxed and shown with pfam descriptions.

TGAGTTTTTCATACTTCAGAACAGTGTGTAATATCCCAAATTTTTATAACTATTTTATAA 79114
AATTAATTTTGTTATAATAAGATTTTATATATATATAATTCTGAAAAATAAAGTAAAATC 79054
ATAATAAATCAGATTTTTATGAAACTGTTTTTATGTATTAAATCAGATTTTTTTTTAAAA 78994
GAAATTTCGAAATCAAAATTTTGAAATTAAATCAGATTTTTATCTTTTGAAGAAAAAAAA 78934
AATTGGAATTTGATTTTCAGATTTTTCGTCCAAAACGAAAAATAGAGAGAAAAGAAAATT 78874
CTGAAATTATAATTTGAGTTGGTTTTGGAAAAGGATTAGATTTTTGTAAATACTTATCTT 78814
TTAGTGAAACCCTAGATTTACATATATATATATATACCCCCAAAACACCAAAAAAATTCT 78754
CACGTAATACACTTTCTTCATCTTTTTGGTAAGTTTTAAATTTCAGATCTAAAATCACCA 78694
TTGTTGTGTTTGAAGGTTTCAACAAAAAAAAAAAAAACTTTTAAAATCCGATGACTTGGT 78634
CGGAGTCCGGCGTCGGTTTTCTTCTTCTCTTCTTCTTCTGTTTTCTTCTTTCTTCTTCCT 78574
TCTTCCTTCTTCTTTCTTCTTTCTATTTCTTCCTTCCTCTTTCTTATTTCATTTGGCTTT 78514
GCGTGCGATGCTTCACAGTACCTCCAAGATTTGATGTTTTTAAAATAAAAAAAGAAAAAA 78454
AAAAAAAGAAAAAGGAATTACATGTGCATTAAAGTCCACGAGAGATTAGACTTACTGTAA 78394
CATGAAAGTTGAAAAAAAATTTCAAGAAAAAATTTCCTTTCAATCTGTGACGACGCTGTC 78334
ACGTTTTTTTTCAAGTAAAAGCTTTTTAATTTTTTTTTCTTTTCAACCTTGGTTATTTAA 78274
CTTTGTCAACTATTAAGTATTAACCTTGTAAGCTTTGATTAATTTTTCTTTCGGAATATT 78214
ATATTAAGTGACTAACATTAGGTAATATTATTTGTTGTAGACGGAAACTTCGTAGAAGGC 78154
GATTTTTAGTTGATATTGCTATACTTGGGACGCGTTTGAGGTAATTCACAATGTCCATCA 78094
ACAACTAGAATCCCGTTAGGAATTGTATGTGCTATGTGTAATGCATGTGGTTATGTGATT 78034
CTATATTTATGTTAGATGTATATTATGTGACAGGAATCGTATAGTACTTTCTATGACTAA 77974
AATTATTATTATTATTATTACTATAGTACTATTATCCCTGCGTATAATATATATGTATCT 77914
ACTGGTATTGTTGTTATGAATTTGGATTTATAATGTACCTAATTATACCATCGTATTAAA 77854
ATTATGCTAATATTGACAATGTACTTAATGGGTATAACTAGTGTGTTAATGATGTAGCGA 77794
ACGGTATTATAAGTTGACATATATTGTAACTCTATGAGGCTCTAATGATGGATATATTGG 77734
TTACTACGAATTATGGACGTTGATTAGTATTAAGTGGCTAAGTTGTGAAAAATATTATTT 77674
GAACTAAAGATGTTCCTGCTAATGTTAATGTGATGTTTGATGTGTCACAACTTTTAAAAA 77614
TCTATTAATCACGTAAAGTGGAAATTGAGGGAATTATCCTGTGGATTTGGATCCTCCATA 77554
GGTGATGAACAGTACTTGATTTTATTATGATACAACTCTTTATTGTCTTCCTCCTAATAC 77494
TATTGGTGCGCATTGCGGATACCCATTAGATTAGTGAGGGGTGTGCACACTAGGGACGCA 77434
CTGCAAAACGATATTGGCCATTGCTCTTAATTTATTGGGTGACGACCCATGTTGAAGTAG 77374
GTGGTTACTGGGATTAGTCCCGGCCTACTGACGTTCGATTCCCCTCTAATAATTAAATGT 77314
GACGTTGTGTGCATCATATAAACTTGGAGTATATAAATTGATTAAGTGTGTTATGTACTT 77254
GTTTGTATTACGTTATACTTGTCTGCATTGTATTATATTTGTATGTATGTTATAAAAATT 77194
ATTTTTAAAATACAACTATTATTATTATTATTTTACGGGATGGAGTTGTAATTACTTAGC 77134
TTTCGCTAATTTTTGTGTTTTTGTTTTTCTTTGCTCTTTTCCTATTTATATTGTGCAGGT 77074
TGGTGAAAGGGACTACGTTGCAGGAATGAGTGAATTTATAGTTTTTAAGTTCACCTAGGT 77014
TAATAAACACTACAGTACTGCCATTTGAGATTGTTTAGTTTTGTTTTGGTATAAAGTAGA 76954
CAACTCAGTTTGGTTTATTTTGATATTTTGGTTTGTGAGTAAAATTTTTATTTCTAATAA 76894
TTAGCCTTGCATTTATTTTGAACAATAGATGCGGCGGTAACACCCTAAAGGTTTGGTCAG 76834
ATATTTTAATGAAAAGTGAATATTTTATAAAGTTTTAGAACCAAAGTTTTGGGGTGTTAC 76774

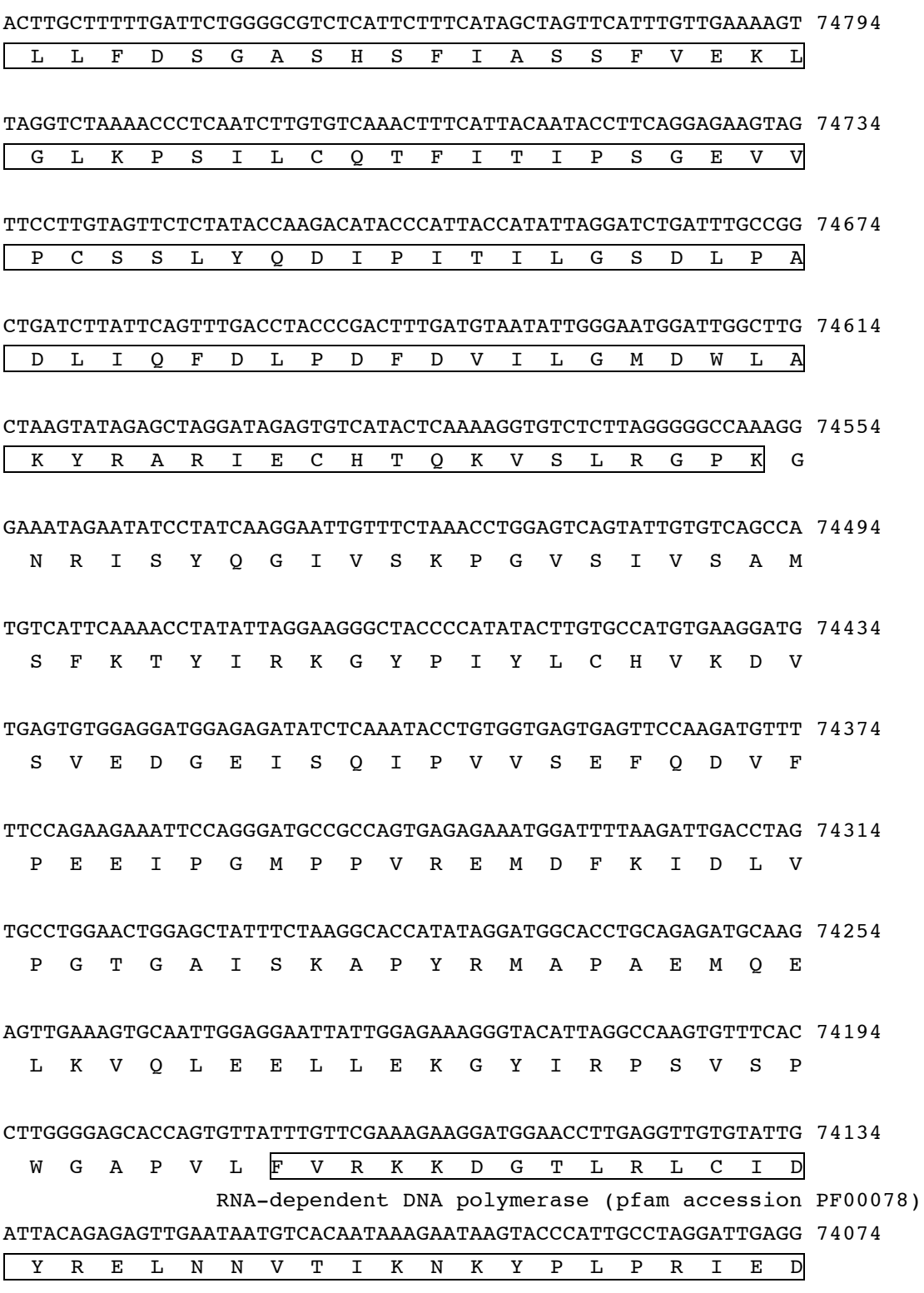
H. Matsuhira et al., FIGURE S4



H. Matsuhira et al., FIGURE S4

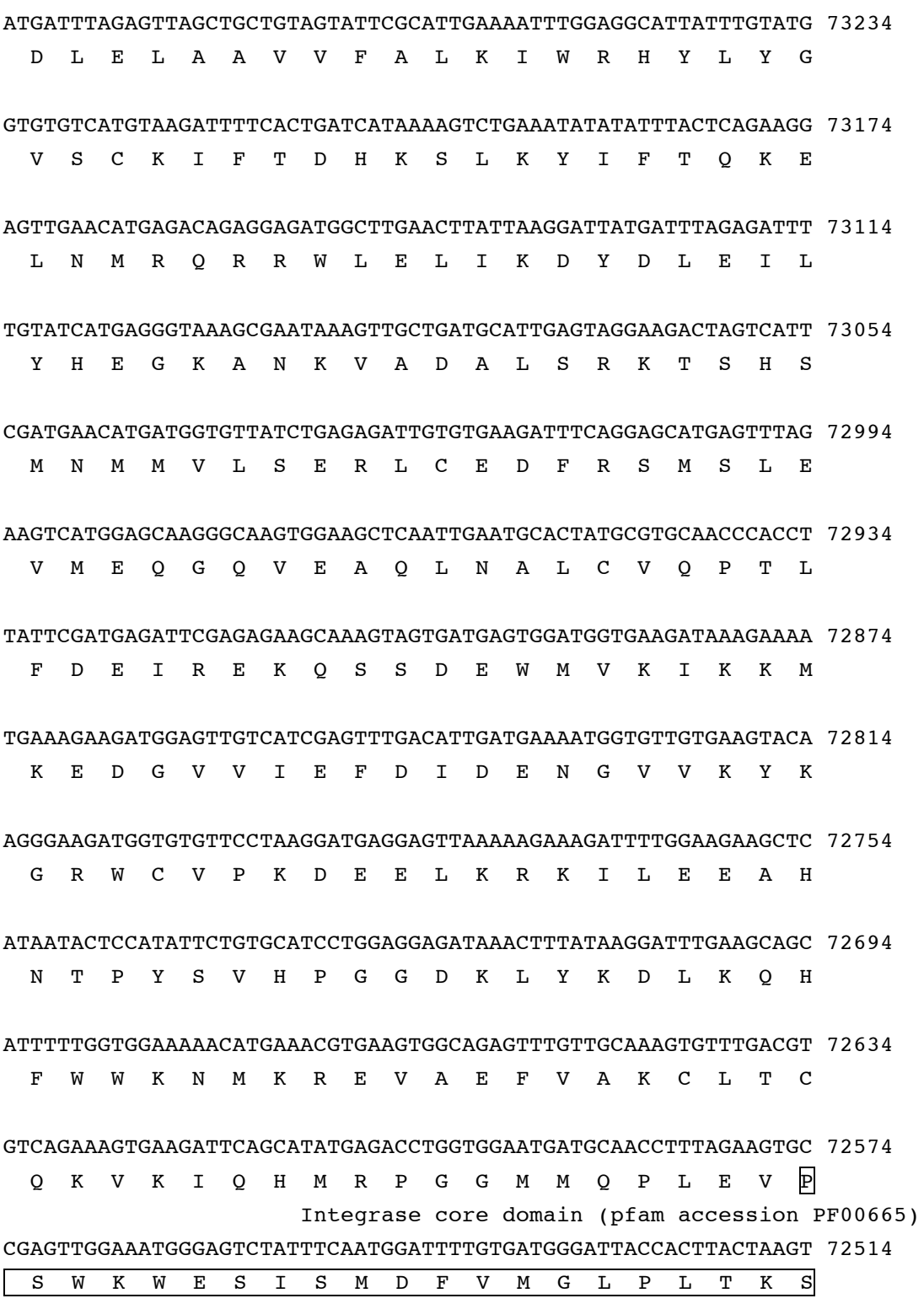


H. Matsuhira et al., FIGURE S4

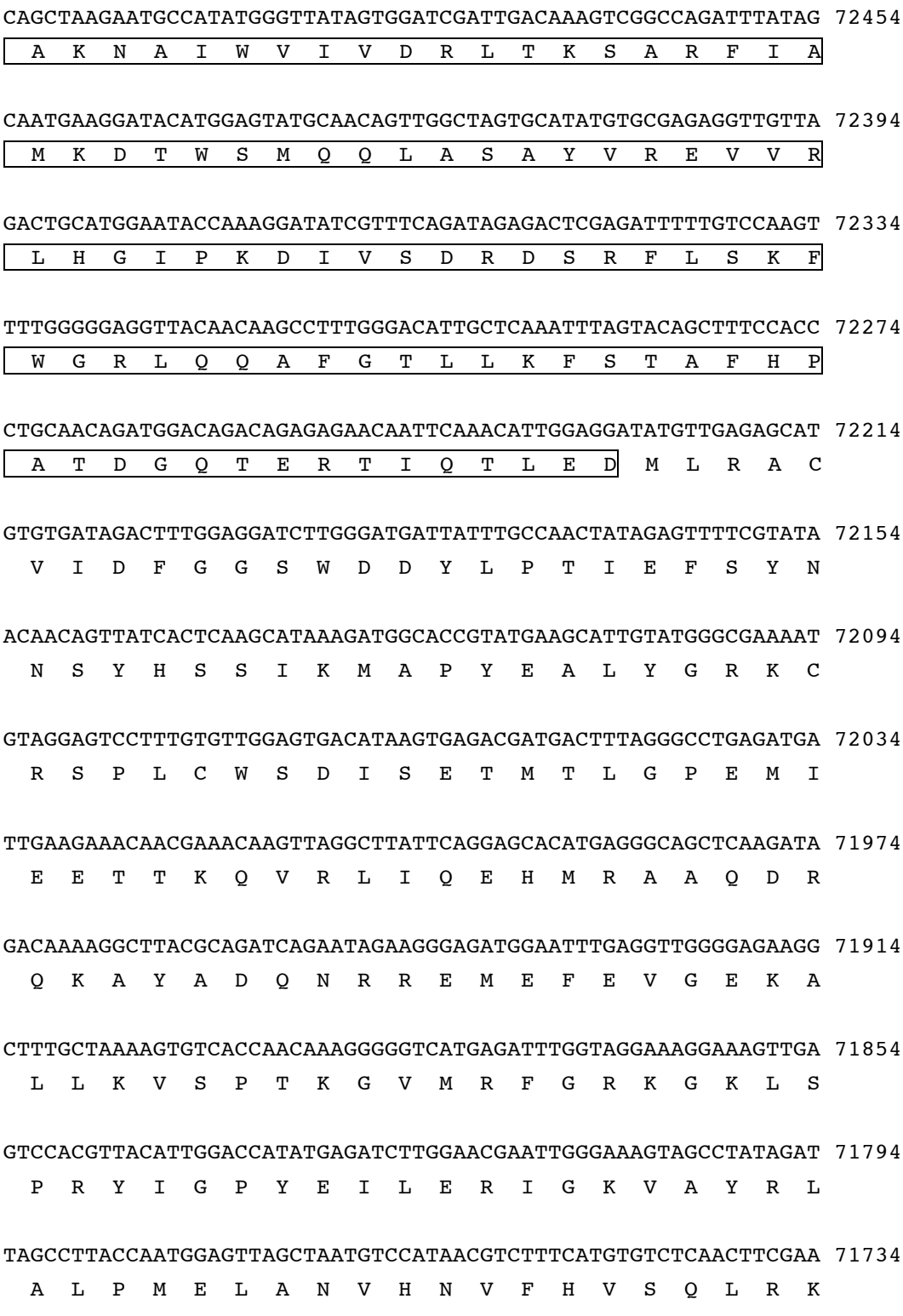


H. Matsuhira et al., FIGURE S4

H. Matsuhira et al., FIGURE S4



H. Matsuhira et al., FIGURE S4



H. Matsuhira et al., FIGURE S4

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

TATCCTTTGAGCAACGCCCAGTTAGGATTCTTGATACCAAAACGAGAAGTACCCGGAACA 71614
  S  F  E  Q  R  P  V  R  I  L  D  T  K  T  R  S  T  R  N  K
           Chromo domain (pfam accession PF00385)
AGGCGGTAAAACTAGTCAAGGTGTTATGGTCAAGTCAAACTTCTGAAGAGGCTACTTGGG 71554
  A  V  K  L  V  K  V  L  W  S  S  Q  T  S  E  E  A  T  W  E

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

TCGGGGACGAAACTCTTTAAGGGGGGTAGAATGTGATACTAACTTTTTGTTTGTATATTA 71434
  G  D  E  T  L  *

GTAGCGAGCGATAACGTTAAAGTTCGAGGACGAACTTTCTTTTAAGGGAGAGTAGATGTA 71374
ATATCCCAAATTTTTATAACTATTTTATAAAATTAATTTTGTTATAATAAGATTTTATAT 71314
ATATATAATTCTGAAAAATAAAGTAAAATCATAATAAATCAGATTTTTATGAAACTGTTT 71254
TTATGTATTAAATCAGATTTTTTTTTAAAAGAAATTTCGAAATCAAAATTTTGAAATTAA 71194
ATCAGATTTTTATCTTTTGAAGAAAAAAAAAAATTGGAATTTGATTTTCAGATTTTTCGTC 71134
CAAAACGAAAAATAGAGAGAAAAGAAAATTCTGAAATTATAATTTGAGTTGGTTTTGGAA 71074
AAGGATTAGATTTTTGTAAATACTTATCTTTTAGTGAAACCCTAGATTTACATATATATA 71014
TATATACCCCCAAAACACCAAAAAAATTCTCACGTAATACACTTTCTTCATCTTTTTGGT 70954
AAGTTTTAAATTTCAGATCTAAAATCACCATTGTTGTGTTTGAAGGTTTCAACAAAAAAA 70894
AAAAAAACTTTTAAAATCCGATGACTTGGTCGGAGTCCGGCGTCGGTTTTCTTCTTCTCT 70834
TCTTCTTCTGTTTTCTTCTTTCTTCTTCCTTCTTCCTTCTTCTTTCTTCTTTCTATTTCT 70774
TCCTTCCTCTTTCTTATTTCATTTGGCTTTGCGTGCGATGCTTCACAGTACCTCCAAGAT 70714
TTGATGTTTTTAAAATAAAAAAAGAAAAAAAAAAAAAGAAAAAGGAATTACATGTGCATT 70654
AAAGTCCACGAGAGATTAGACTTACTGTAACATGAAAGTTGAAAAAAAATTTCAAGAAAA 70594
AATTTCCTTTCAATCTGTGACGACGCTGTCACGTTTTTTTTCAAGTAAAAGCTTTTTAAT 70534
TTTTTTTTCTTTTCAACCTTGGTTATTTAACTTTGTCAACTATTAAGTATTAACCTTGTA 70474
AGCTTTGATTAATTTTTCTTTCGGAATATTATATTAAGTGACTAACATTAGGTAATATTA 70414
TTTGTTGTAGACGGAAACTTCGTAGAAGGCGATTTTTAGTTGATATTGCTATACTTGGGA 70354
CGCGTTTGAGGTAATTCACAATGTCCATCAACAACTAGAATCCCGTTAGGAATTGTATGT 70294
GCTATGTGTAATGCATGTGGTTATGTGATTCTATATTTATGTTAGATGTATATTATGTGA 70234
CAGGAATCGTATAGTACTTTCTATGACTAAAATTATTATTATTATTATTACTATAGTACT 70174
ATTATCCCTGCGTATAATATATATGTATCTACTGGTATTGTTGTTATGAATTTGGATTTA 70114
TAATGTACCTAATTATACCATCGTATTAAAATTATGCTAATATTGACAATGTACTTAATG 70054
GGTATAACTAGTGTGTTAATGATGTAGCGAACGGTATTATAAGTTGACATATATTGTAAC 69994
TCTATGAGGCTCTAATGATGGATATATTGGTTACTACGAATTATGGACGTTGATTAGTAT 69934
```


H. Matsuhira et al., FIGURE S4

TAAGTGGCTAAGTTGTGAAAAATATTATTTGAACTAAAGATGTTCCTGCTAATGTTAATG 69874
TGATGTTTGATGTGTCACAACTTTTAAAAATCTATTAATCACGTAAAGTGGAAATTGAGG 69814
GAATTATCCTGTGGATTTGGATCCTCCATAGGTGATGAACAGTACTTGATTTTATTATGA 69754
TACAACTCTTTATTGTCTTCCTCCTAATACTATTGGTGCGCATTGCGGATACCCATTAGA 69694

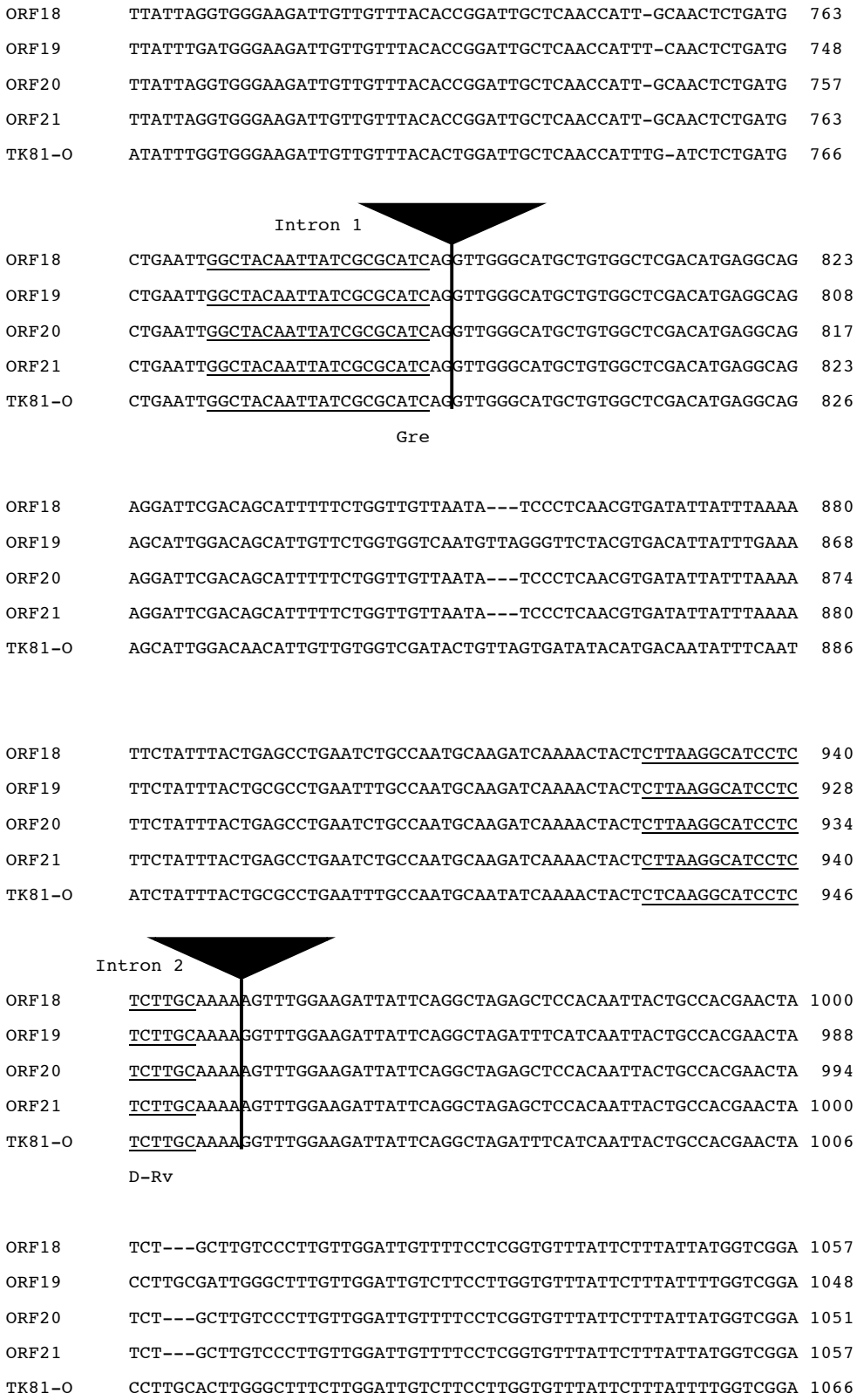
FIGURE S4.– Nucleotide sequence of *bvgypsy-1*. Numbers of nucleotide residues are coordinated with those of BAC clone 33E19. The open reading frame is indicated with its putative translation product. The 1684-bp repeated sequences are underlined. Putative domains are boxed and shown with pfam descriptions.

```
ORF18      ATGGCGTGGTACAGAAATTCAAGGTTTGTCTACAATGCTTTAAAACTCAACTTGCGTTCC   60
ORF19      ATGGCGTGGTACAGAAATTCAAGGTTTGTCTACAATGCTTTAAAACTCAACTTGCGTTCC   60
ORF20      ATGGCGTGGTACAGAAATTCAAGGTTTGTCTACAATGCTTTAAAACTCAACTTGCGTTCC   60
ORF21      ATGGCGTGGTACAGAAATTCAAGGTTTGTCTACAATGCTTTAAAACTCAACTTGCGTTCC   60
TK81-O     ATGGCGTGGTACAGAAATTCAAGGTTTGTCTACAATGCTTTAAAACTCAACTTGCGTTCC   60

ORF18      AAAACATTTGGTACTATTCCAACTCCAAGAGTTCATTCGAATTCCTCATCTTTGTTTTAC  120
ORF19      AAAACATTTGGTACTATTCCAACTCCAAGAGTTCATTCGAATTCCTCATCTTTGTTTTAC  120
ORF20      AAAACATTTGGTACTATTCCAACTCCAAGAGTTCATTCGAATTCCTCATCTTTGTTTTAC  120
ORF21      AAAACATTTGGTACTATTCCAACTCCAAGAGTTCATTCGAATTCCTCATCTTTGTTTTAC  120
TK81-O     AAAACATTTGGTACTATTCCAACTCCAAGAGTTCATTCGAATTCCTCATCTTTGTTTTAC  120

ORF18      AATCAATCTACTAATAAGTGTAGTGGGTTATTTGGGTCTGCAAAATCTGGGTATTTTAAT  180
ORF19      AATCAATCTACTAATAAGTGTAGTGGGTTATTTGGGTCTGCAAAATCTGGGTATTTTAAT  180
ORF20      AATCAATCTACTAATAAGTGTAGTGGGTTATTTGGGTCTGCAAAATCTGGGTATTTTAAT  180
ORF21      AATCAATCTACTAATAAGTGTAGTGGGTTATTTGGGTCTGCAAAATCTGGGTATTTTAAT  180
TK81-O     AATCAATCTACTAA---GTGTAGTGGGTTATTTGGGTCTGCAAAATCTGGGTATTTTAAT  177

ORF18      GGGTTTAAACATCATCAAGAGATTAGCTCTTTCTCTGGTTTTGCAAGGAGAAATTATCAT  240
ORF19      GGGTTTAAACATCATCAAGAGATTAGCTCTTTCTCTGGTTTTGCAAGGAGAAATTATCAT  240
ORF20      GGGTTTAAACATCATCAAGAGATTAGCTCTTTCTCTGGTTTTGCAAGGAGAAATTATCAT  240
ORF21      GGGTTTAAACATCATCAAGAGATTAGCTCTTTCTCTGGTTTTGCAAGGAGAAATTATCAT  240
TK81-O     GGGTTTAAACATCATCAAGAGATTAGCTCTTTCTCTGGTTTTGCAAGGAGAAACTATCAT  237

ORF18      GGTGATAAAACCGAAGTAAGTGTTGAATCATGGCTGGAAAAATTCCTTGTTCCAATTGGA  300
ORF19      GGTGATAAAACCGAAGTAAGTGCTGAATCATTGCTGGAAAAATTACTTCTTC---TTGCA  297
ORF20      GGTGATAAAACCGAAGTAAGTGTTGAATCATGGCTGGAAAAATTCCTTGTTCCAATTGGA  300
ORF21      GGTGATAAAACCGAAGTAAGTGTTGAATCATGGCTGGAAAAATTCCTTGTTCCAATTGGA  300
TK81-O     GGTGTTAAAACCGAAGTAAGTGTTGAATTTCGGGTGGAAAAATTACTTCTTGGAATTGCA  297

ORF18      C---------TAATCTTGACTTTTGGTATACTTGGTTACCCTCATGTGCACCCAGTAGTT  351
ORF19      GTTGCAC---TAATCTTGA------------TTGCTTACCGTCATGTGCACCCAGTAGTT  342
ORF20      C---------TAATCTTGACTTTTGGTATACTTGGTTACCCTCATGTGCACCCAGTAGTT  351
ORF21      C---------TAATCTTGACTTTTGGTATACTTGGTTACCCTCATGTGCACCCAGTAGTT  351
TK81-O     C------TAATAATCTCGCATTCTGGTATGATTGCTTTCTTTTATTTGCACCCAGTAGTT  351
```


H. Matsuhira et al., FIGURE S5

```
ORF18      GTGCCATATACAGGAAGGAAGCATTATGTGCTTATGTCAACAACTCGTGAGAATGAAATT  411
ORF19      GTGCCATATACAGGAAGGAAGCATTATGTGCTTATGTCAACAACTCGTGAGAATGAAAAT  402
ORF20      GTGCCATATACAGGAAGGAAGCATTATGTGCTTATGTCAACAACTCGTGAGAATGAAATT  411
ORF21      GTGCCATATACAGGAAGGAAGCATTATGTGCTTATGTCAACAACTCGTGAGAATGAAATT  411
TK81-O     GTGCCATATACAGGAAGGAAGCATTATGTGATTTTGTCAACAACTCATGAGAATGAAAAT  411
           D-Fw

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

ORF18      TCAATATTCCAACACATTCTTGAATCACTGGAAAGAGAGATTAATCACCATGAACTCGAA  531
ORF19      TCTATATTCCAACACATTATTGAATCACTGGAAAGAGAGATTAATCACCATGAACTCGAA  522
ORF20      TCAATATTCCAACACATTCTTGAATCACTGGAAAGAGAGATTAATCACCATGAACTCGAA  531
ORF21      TCAATATTCCAACACATTCTTGAATCACTGGAAAGAGAGATTAATCACCATGAACTCGAA  531
TK81-O     TCTATATTCCAACACATTCTTGAATCACTGGAAAGAGAGATTAATCACCATGAACTCGAA  531

ORF18      CTCGAACTCGAA------AGAGATGAAACTTTCAAGGAGAAAACCATTTGGAAGGAGGAG  585
ORF19      CTCGAA------------AGAGATGAAACTTTCAAGGAGAAAACCATTTGGAAGGAGGAG  570
ORF20      CTCGAA------------AGAGATGAAACTTTCAAGGAGAAAACCATTTGGAAGGAGGAG  579
ORF21      CTCGAACTCGAA------AGAGATGAAACTTTCAAGGAGAAAACCATTTGGAAGGAGGAG  585
TK81-O     CTCGAACTCGAACTCGAAAGAGATGAAACTTTCAAGGAGAAAACCATTTGGAAGGAGGAG  591

ORF18      ACAGTTGATGATAAAGATAGTAGGAAGAAGCATAGTGGGGCTAAGATAACTACTAACCAT  645
ORF19      ACAGTTGATGATAAAGATAGTAGGAAGAAGCATAGTGGGGCTAAGATAACTACTAACCAT  630
ORF20      ACAGTTGATGATAAAGATAGTAGGAAGAAGCATAGTGGGGCTAAGATAACTACTAACCAT  639
ORF21      ACAGTTGATGATAAAGATAGTAGGAAGAAGCATAGTGGGGCTAAGATAACTACTAACCAT  645
TK81-O     ACAGATCATGATAAAGATAGTAGGAAGAAGCATAGTGGGGCTAAGATAACTACTAACCAT  651

ORF18      TTGGAAGGGATGAATTGGGAAATTTTCGTTGTTGATAAACCGTTGGTTGAGTCCAGTTAT  705
ORF19      TTGGAAGGGTTGAATTGGGAAATTTTCGTTGTTGATAAACCGTTGGTTGAGTCCAGTTGT  690
ORF20      TTGGAAGGGATGAATTGGGAAATTTTCGTTGTTGATAAACCGTTGGTTGAGTCCAGTTAT  699
ORF21      TTGGAAGGGATGAATTGGGAAATTTTCGTTGTTGATAAACCGTTGGTTGAGTCCAGTTAT  705
TK81-O     ---GAAGGGATGAATTGGGAAATTTTCGTTGTCGATAAACCGTGGGTTGAGTCCAGTTGT  708
```


H. Matsuhira et al., FIGURE S5



H. Matsuhira et al., FIGURE S5

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

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

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

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

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

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

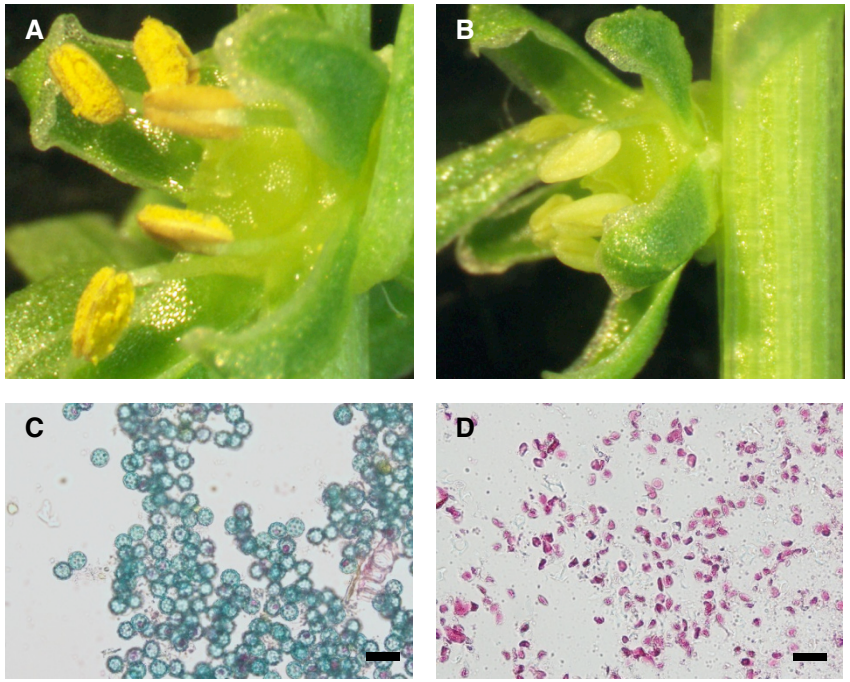


FIGURE S7.– Comparison of anther morphology and anther content between two sugar-beet plants derived from the 14 F1 plants. Panels A and C show photographs taken from a plant having the biaphos-resistance gene. Panels B and D show photographs taken from a plant missing the bialaphos-resistance gene.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

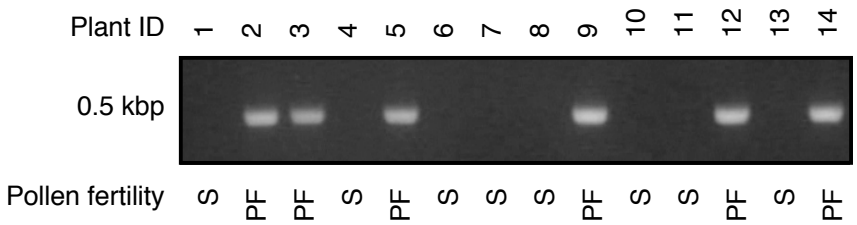


FIGURE S8.– Co-segregation analysis between the bialaphos-resistance gene and partial fertility. Agarose gel electrophoresis of PCR products resulting from amplification using primers targeting the bialaphos-resistance gene. Size of the amplicon is shown on the left. Plant ID and pollen fertility is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

```
bvORF18/21 MAWYRNSRFVYNALKLNLRSKTFGTIPTPRVHSNSSSLFYNQSTNKCSGLFGSAKSGYFN 60
bvORF19    MAWYRNSRFVYNALKLNLRSKTFGTIPTPRVHSNSSSLFYNQSTNKCSGLFGSAKSGYFN 60
bvORF20    MAWYRNSRFVYNALKLNLRSKTFGTIPTPRVHSNSSSLFYNQSTNKCSGLFGSAKSGYFN 60
bvORF20L   MAWYRNSRFVYNALKLNLRSKTFGTIPTPRVHSNSSSLFYNQST-KCSGLFGSAKSGYFN 59
           ************************************************ ***************

bvORF18/21 GFKHHQEISSFSGFARRNYHGDKTEVSVESWLEKFLVPIGLILTFG-ILGYPHVHPVVVP 119
bvORF19    GFKHHQEISSFSGFARRNYHGDKTEVSAESLLEKLLL---LAVALI-LIAYRHVHPVVVP 116
bvORF20    GFKHHQEISSFSGFARRNYHGDKTEVSVESWLEKFLVPIGLILTFG-ILGYPHVHPVVVP 119
bvORF20L   GFKHHQEISSFSGFARRNYHGVKTEVSVEFRVEKLLLGIALIISHSGMIAFFYLHPVVVP 119
           ********************** *****.*  :**:*:   * ::   ::.: ::******

bvORF18/21 YTGRKHYVLMSTTRENEIGEVEKRKIQPATHPDTDRVRSIFQHILESLEREINHHELELE 179
bvORF19    YTGRKHYVLMSTTRENENGEVEKRKIQPATHPDTERVRSIFQHIIESLEREINHHELELE 176
bvORF20    YTGRKHYVLMSTTRENEIGEVEKRKIQPATHPDTDRVRSIFQHILESLEREINHHELELE 179
bvORF20L   YTGRKHYVILSTTHENENGEFEKRKIQPATHPDTERVRSIFQHILESLEREINHHELELE 179
           ********::***:*** **.*************:*********:***************

bvORF18/21 LE--RDETFKEKTIWKEETVDDKDSRKKHSGAKITTNHLEGMNWEIFVVDKPLVESSYLL 237
bvORF19    ----RDETFKEKTIWKEETVDDKDSRKKHSGAKITTNHLEGLNWEIFVVDKPLVESSCLF 232
bvORF20    ----RDETFKEKTIWKEETVDDKDSRKKHSGAKITTNHLEGMNWEIFVVDKPLVESSYLL 235
bvORF20L   LELERDETFKEKTIWKEETDHDKDSRKKHSGAKITTNH-EGMNWEIFVVDKPWVESSCIF 238
               *************** .****************** **:********** **** ::

bvORF18/21 GGKIVVYTGLLNHCNSDAELATIIAHQVGHAVARHEAEDSTAFFWL-LISLNVILFKILF 296
bvORF19    DGKIVVYTGLLNHFNSDAELATIIAHQVGHAVARHEAEHWTALFWWSMLGFYVTLFEILF 292
bvORF20    GGKIVVYTGLLNHCNSDAELATIIAHQVGHAVARHEAEDSTAFFWL-LISLNVILFKILF 294
bvORF20L   GGKIVVYTGLLNHCISDAELATIIAHQVGHAVARHEAEHWTTLLWSILLVIYMTIFQYLF 298
           .************  ***********************. *:::*  :: : : :*: **

bvORF18/21 TEPESANARSKLLLRHPLLQKVWKIIQARAPQLLPR-TICLSLVGLFSSVFILYYGRKEI 355
bvORF19    TAPEFANARSKLLLRHPLLQKVWKIIQARFHQLLPRTTLRLGFVGLSSLVFILYFGRKEI 352
bvORF20    TEPESANARSKLLLRHPLLQKVWKIIQARAPQLLPR-TICLSLVGLFSSVFILYYGRKEI 353
bvORF20L   TAPEFANAISKLLSRHPLLQKVWKIIQARFHQLLPRTTLHLGFLGLSSLVFILYFGRKEI 358
           * ** *** **** ***************  ***** *: *.::** * *****:*****


bvORF18/21 EADHIGVLLMASAGYDPRVAPQVYDKLAKPLGDWNCLATHPFARMRAKLLARADVMKEAD 415
bvORF19    EADHIGVLLMASAGYDPRVAPQVYDKLAKPLGDWNCLATHPFARMRAKLLARADVMKEAD 412
bvORF20    EADHIGVLLMASAGYDPRVAPQVYDKLAKPLGDWNCLATHPFARMRAKLLARADVMKEAD 413
bvORF20L   EADHIGVLLMASAGYDPRVAPQVYDKLAKPLGDWNCLATHPFARMRAKLLARADVMKEAD 418
           ************************************************************

bvORF18/21 KIYNEVVAGRAIQGLQ 431
bvORF19    KIYNEVVAGRAIQGLQ 428
bvORF20    KIYNEVVAGRAIQGLQ 429
bvORF20L   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 for phylogenetic analysis.

Name of genes	Label in tree
AtPPR_1g01970[*1]	AT1g01970
AtPPR_1g02060	AT1g02060
AtPPR_1g02150	AT1g02150
AtPPR_1g02370	AT1g02370
AtPPR_1g02420	AT1g02420
AtPPR_1g03100	AT1g03100
AtPPR_1g03560	AT1g03560
AtPPR_1g05600	AT1g05600
AtPPR_1g05670	AT1g05670
AtPPR_1g06270	AT1g06270
AtPPR_1g06580	AtRFL1
AtPPR_1g06710	AT1g06710
AtPPR_1g07590	AT1g07590
AtPPR_1g07740	AT1g07740
AtPPR_1g08610	AT1g08610
AtPPR_1g09680	AT1g09680
AtPPR_1g09820	AT1g09820
AtPPR_1g09900	AT1g09900
AtPPR_1g10270	AT1g10270
AtPPR_1g10910	AT1g10910
AtPPR_1g11630	AT1g11630
AtPPR_1g11710	AT1g11710
AtPPR_1g11900	AT1g11900
AtPPR_1g12300	AtRFL2
AtPPR_1g12620	AtRFL3
AtPPR_1g12700	AtRFL4
AtPPR_1g12770	AT1g12770
AtPPR_1g13040	AT1g13040
AtPPR_1g13630	AT1g13630
AtPPR_1g13800	AT1g13800
AtPPR_1g15480	AT1g15480
AtPPR_1g16830	AT1g16830
AtPPR_1g18900	AT1g18900
AtPPR_1g19290	AT1g19290
AtPPR_1g19520	AT1g19520
AtPPR_1g20300	AT1g20300
AtPPR_1g22960	AT1g22960
AtPPR_1g26460	AT1g26460
AtPPR_1g26500	AT1g26500
AtPPR_1g28020	AT1g28020
AtPPR_1g30610	AT1g30610
AtPPR_1g31790	AT1g31790
AtPPR_1g31840	AT1g31840
AtPPR_1g43010	AT1g43010
AtPPR_1g51965	AT1g51965
AtPPR_1g52620	AT1g52620
AtPPR_1g52640	AT1g52640
AtPPR_1g53330	AT1g53330
AtPPR_1g55630	AT1g55630
AtPPR_1g55890	AT1g55890
AtPPR_1g60770	AT1g60770
AtPPR_1g61870	AT1g61870

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

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

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

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

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

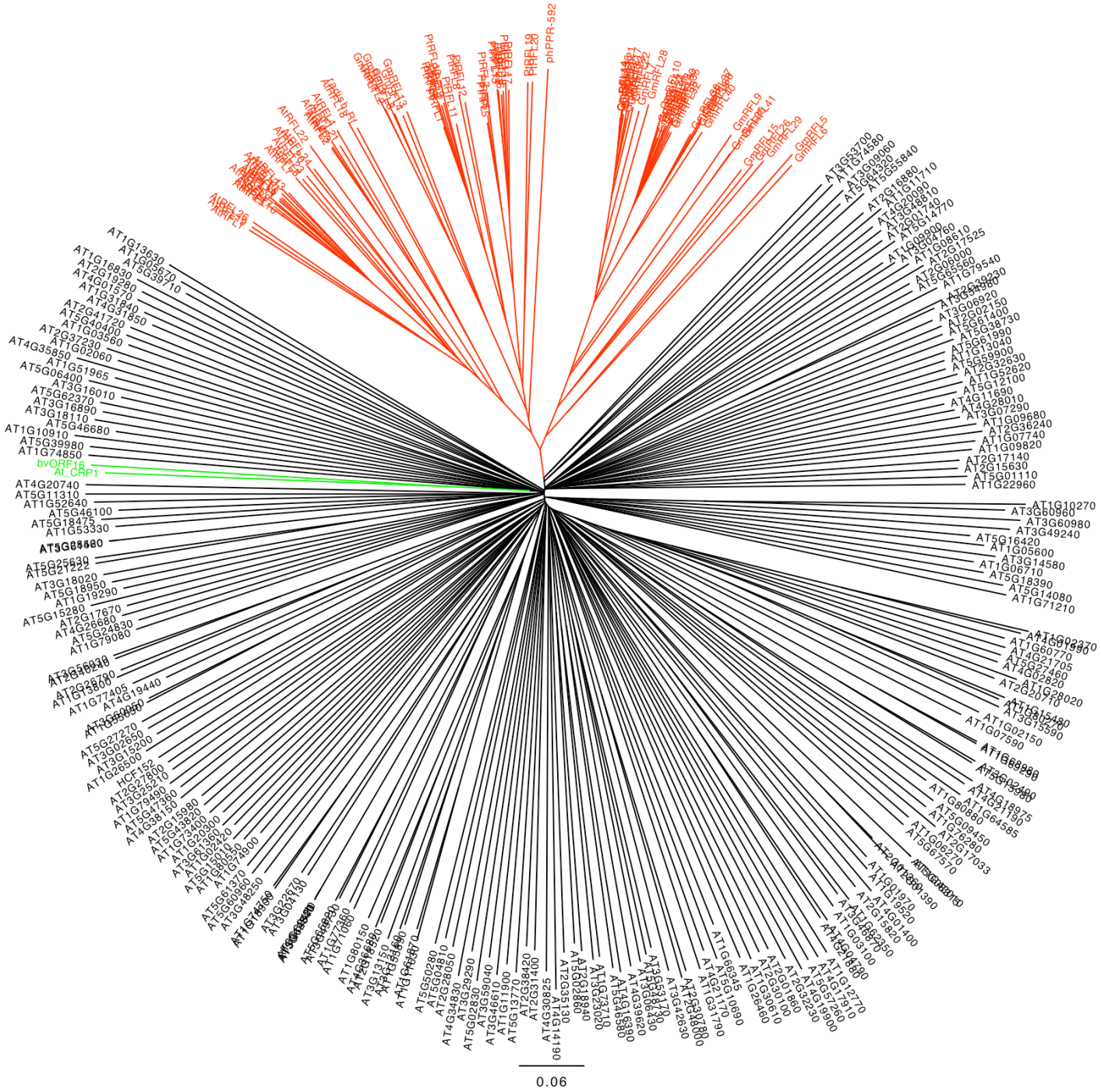
[*1]http://www.plantenergy.uwa.edu.au/applications/ppr/ppr.php and http://www.arabidopsis.org/tools/bulk/sequences/index.jsp

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

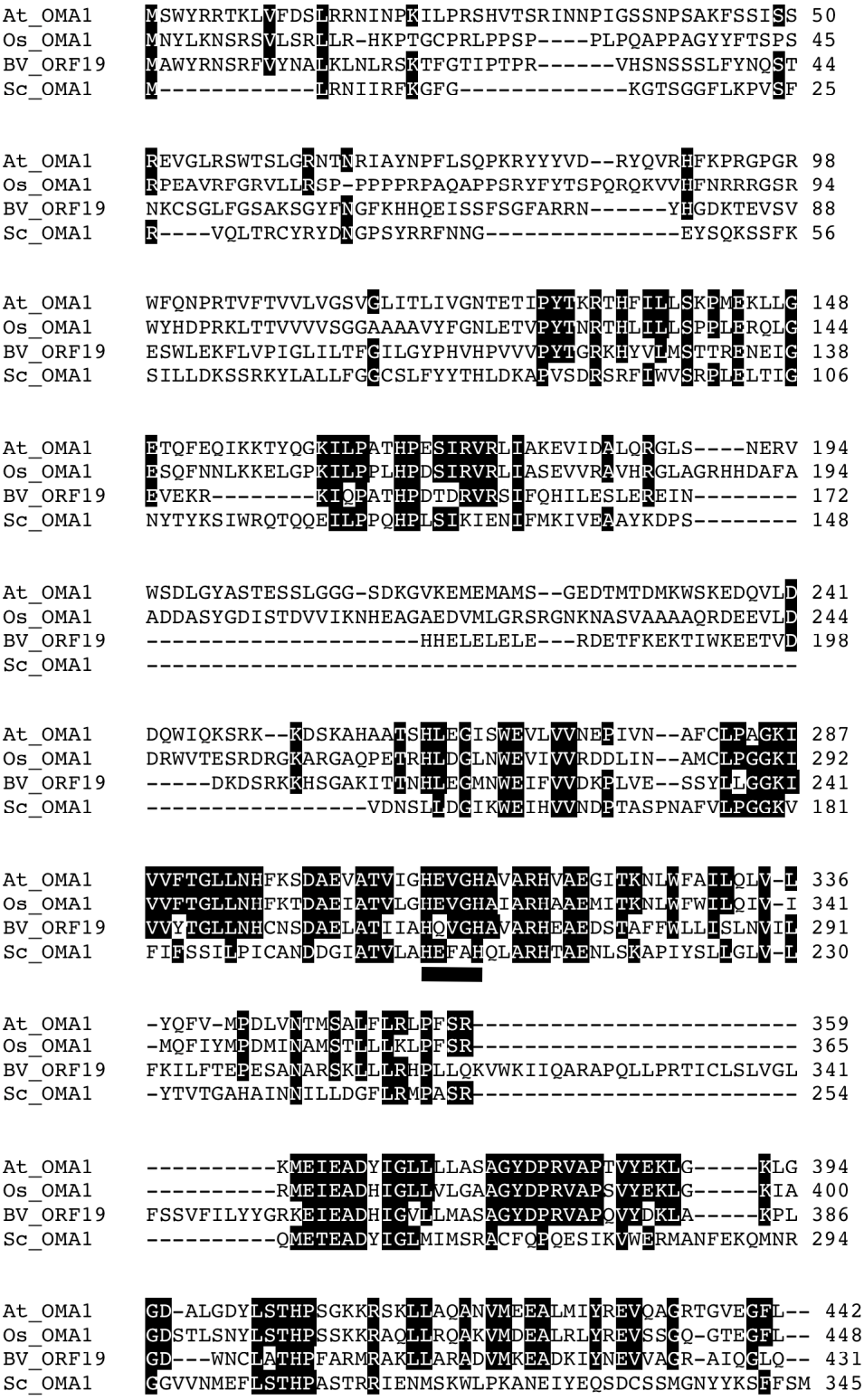
[*3]http://www.phytozome.net/search.php?show=blast&method=Org_Ptrichocarpa_v2.2

[*4]DDBJ/GenBank/EMBL dataase

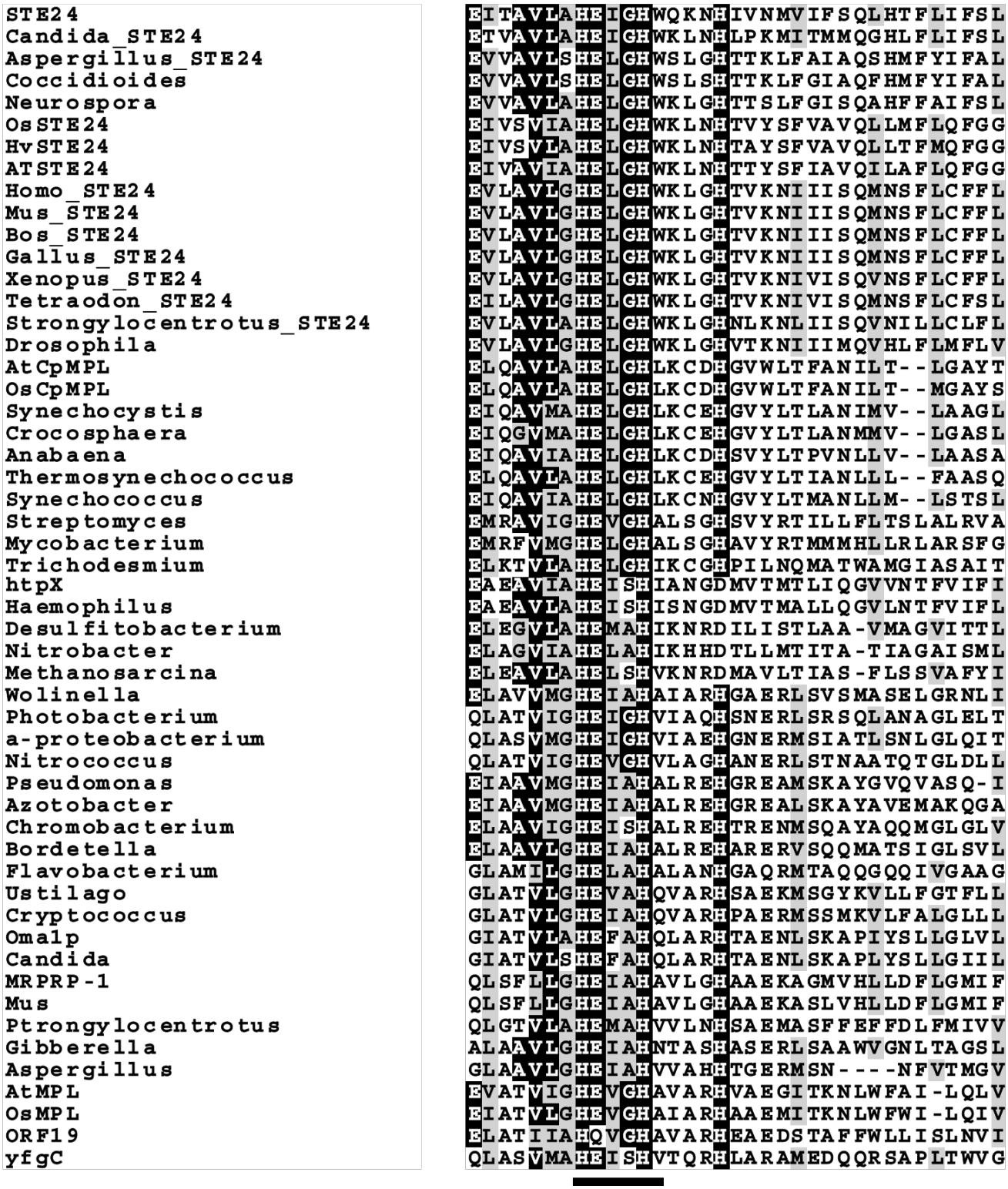
[*5]This study



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

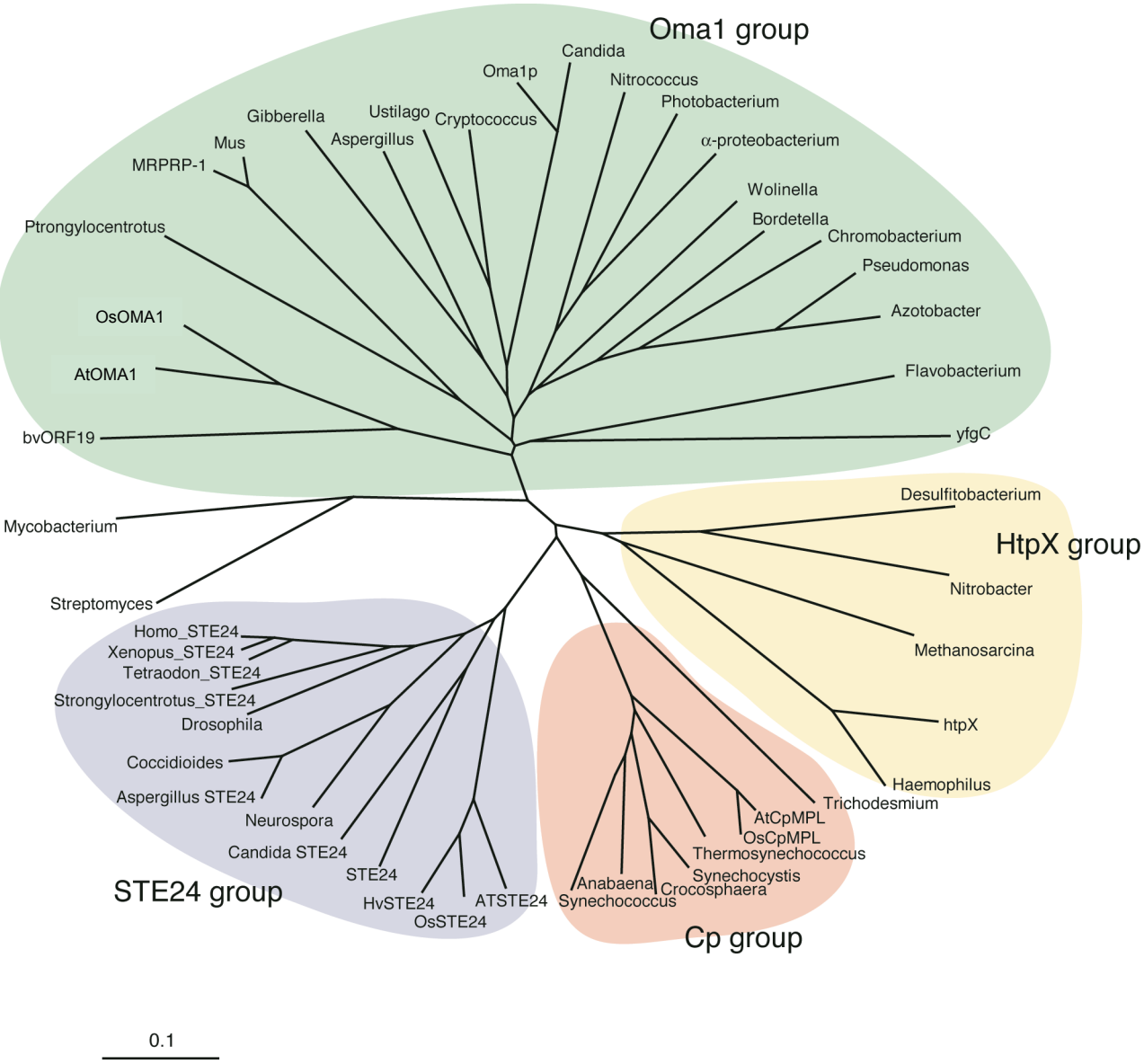


File S3. Multiple alignment of amino acid sequences of OMA1-homologous proteins from Arabidopsis (At_OMA1, At5g51740), rice (Os_OMA1, Os02g0735100), sugar beet (bvORF19, this study), and yeast (Sc_OMA1, S000001795). Position of the Zn^{2+} binding motif is shown by a horizontal line. The amino acid sequences 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.



File S4. See next page for the legend.

File S4. Multiple alignment of ~35 amino acid residues surrounding the Zn^{2+} binding motif of peptidase M48 proteins, a protein family to which yeast OMA1 belongs. The position of the Zn^{2+} 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; a-proteobacterium, a-proteobacterium HTCC2255, ZP_01448796; Nitrococcus, *Nitrococcus mobilis,* ZP_01126393; Pseudomonas, *Pseudomonas aeruginosa,* NP_253322; Azotobacter, *Azotobacter vinelandii,* ZP_00416091; Chromobacterium, *Chromobacterium violaceum,* NP_899823; Bordetella, *Bordetella bronchiseptica,* NP_888655; Flavobacterium, *Flavobacterium* sp. MED217, ZP_01061128; Ustilago, *Ustilago maydis,* XP_757961; Cryptococcus, *Cryptococcus neoformans,* XP_569916; Oma1p, *Saccharomyces cerevisiae,* P36163; Candida, *Candida glabrata,* XP_446463; MPRP-1, *Homo sapiens,* BAC79381; Mus, *Mus musculus,* NP_080185; Strongylocentrotus, *Strongylocentrotus purpuratus,* XP_799173; Gibberella, *Gibberella zeae,* XP_390368; Aspergillus, *Aspergillus nidulans,* XP_659454; AtMPL, *Arabidopsis thaliana,* At5g51740 (AtOMA1); OsMPL, *Oryza sativa,*Os02g0735100 (OsOMA1) . Multiple alignment was done by using ClustalX (http://www.clustal.org/clustal2/).



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