The active miniature inverted-repeat transposable element *mPing* post-transcriptionally produces new transcriptional variants in the rice genome

Rise Kum[1], Takuji Tsukiyama[1,*], Haruka Inagaki[1], Hiroki Saito[1], Masayoshi Teraishi[1], Yutaka Okumoto[1], Takatoshi Tanisaka[1,2]

[1]Division of Agronomy and Horticulture Science, Graduate School of Agriculture, Kyoto University, Kyoto 606-8502, Japan

[2]Department of Agriculture for Regional Reclamation, Kibi International University, Minami-Awaji 656-0484, Japan

[*]Corresponding author.

Takuji Tsukiyama,

Division of Agronomy and Horticulture Science, Graduate School of Agriculture, Kyoto University, Kitashirakawa, Sakyoku, Kyoto 606-8502, Japan

E-mail: takuji@kais.kyoto-u.ac.jp, Tel: +81-75-753-6046, Fax +81-75-753-6047

Abstract

Post-transcriptional RNA processing inclusive of alternative splicing and alternative polyadenylation, as well as transcriptional regulation, plays important regulatory roles in eukaryotic gene expression. In eukaryotic genomes, transposable elements (TEs) can alter gene expression at both transcriptional and post-transcriptional levels. *Miniature Ping* (*mPing*) is an active miniature inverted-repeat TE discovered in the rice genome, and its insertion renders adjacent genes stress-inducible. In this study, we examined the effect of *mPing* insertion into coding sequences on RNA processing. The 3' RACE (rapid amplification of cDNA ends) analysis of mutant alleles, each harboring an *mPing* insertion, revealed that *mPing* induced various alternative splicing events. Furthermore, it was found that *mPing* induced alternative polyadenylation within its sequence. In the mutant allele, the body region of *mPing* was heavily methylated, whereas the *mPing*-flanking regions were moderately methylated. These results indicate that *mPing* alters transcript structures post-transcriptionally via induction of alternative splicing that most likely depends on DNA methylation. Based on these results, we discuss the availability of *mPing* as an insertional mutagen in rice.

Keywords: Rice; Transposable element; *mPing*; Alternative splicing; Alternative polyadenylation; Post-transcriptional regulation

Introduction

In eukaryotic genomes, the expression of genes is controlled by transcriptional and post-transcriptional regulatory mechanisms. Alternative splicing is one of the post-transcriptional regulatory mechanism widely adopted in multicellular organisms (Nilsen and Graveley 2010; Kornblihtt et al. 2013). Genome-wide analyses show that a large fraction of the protein-coding genes of multicellular organisms are alternatively spliced, whereas no such alternative splicing has been detected in unicellular organisms (Ast 2004). In human, approximately 95% of genes are alternatively spliced (Pan et al. 2008). Also in plants, more than 60% of genes containing introns undergo alternative splicing (Marquez et al. 2012; Syed et al. 2012). Other than alternative splicing, alternative polyadenylation is important for regulating gene expression in both animals and plants (Mayr and Bartel 2009; Mangone et al. 2010; Xing and Li 2011). In Arabidopsis and rice, 70% and 50% of genes have at least one polyadenylation site with microheterogeneity, respectively (Shen et al. 2008; Wu et al. 2011).

Transposable elements (TEs) are DNA fragments that can move from the original position to any position in the genome. TEs had been thought to be selfish elements for a long time since McClintock (1950) first discovered them through analyzing unstable phenotypes of maize kernels. The progress of genome projects in various organisms, however, revealed that most eukaryotic genomes consist of large numbers of different types of TEs; 35% and over 85% of rice (*Oryza*

56 *sativa*) and maize genomes consist of TEs, respectively (Turcotte et al. 2001; Schnable et al. 2009).
57 Recently, TEs have been recognized to be a major player in genomic evolution by causing genome
58 rearrangements and by altering the structure and regulation of individual genes (Feschotte and
59 Pritham 2007). Furthermore, it has been proposed that TEs contribute to the evolution of regulatory
60 network by altering gene expression at both transcriptional and post-transcriptional levels (Feschotte
61 2008).

62 Miniature inverted-repeat transposable elements (MITEs) are non-autonomous TEs widely
63 deployed in both prokaryotic and eukaryotic genomes. In the sequenced rice genome (cultivar
64 Nipponbare), MITEs are present in >70,000 copies, and many of them are found in the 5'
65 untranslated regions (UTRs), the 3' UTRs, and in the proximity of genes (Oki et al. 2008). Since the
66 5' and 3' UTRs are known to play important roles in gene expression (Chan and Yu 1998; Cazzola
67 and Skoda 2000; Mazumder et al. 2005; Misquitta et al. 2006; Lytle JR et al. 2007;
68 Aguilar-Hernández and Guzmán 2013), MITEs located in the 5' and 3' UTRs are considered to
69 influence the regulation of gene expressions. Furthermore, for over 300 protein-coding genes in rice,
70 coding sequences, polyadenylation sites, transcription start sites, and splicing sites overlap with
71 MITEs (Oki et al. 2008). These indicate that MITEs have greatly contributed to gene expression not
72 only at the transcriptional level but also at the post-transcriptional level in the evolution of the rice
73 genome.

Miniature *Ping* (*mPing*) is the only active MITE identified in the rice genome (Jiang et al. 2003; Kikuchi et al. 2003; Nakazaki et al. 2003). *mPing* is a 430-bp element including 15-bp terminal inverted repeats (TIRs). Although *mPing* is inactive in most rice cultivars, the transposition of *mPing* is activated by various stress treatments, such as cell culture (Jiang et al. 2003), anther culture (Kikuchi et al. 2003), gamma irradiation (Nakazaki et al. 2003), hydrostatic pressure (Lin et al. 2006), and introgression of closely related genome (Shan et al. 2005). Interestingly, in several *japonica* landraces including a strain EG4 (cultivar Gimbozu), *mPing* is still actively transposing under natural growth conditions (Naito et al. 2006). Recently, it was found that, in EG4, *mPing* is mobilized in the embryo with the aid of the developmental stage-specific up-regulation of its autonomous element, *Ping* (Teramoto et al. 2014). *mPing* preferentially transposes into within 0.5-kb upstream of gene, and renders adjacent genes stress inducible (Naito et al. 2009; Yasuda et al. 2013), which indicates that, like other MITEs, *mPing* also contribute to the generation of new regulatory networks at the transcriptional level. Little is known, however, about the effects of *mPing* on the post-transcriptional regulation of genes. In this study, we demonstrate that *mPing* is creating new transcript isoforms by inducing various alternative splicing events. Furthermore, we discuss the possible mechanisms of alternative splicing induced by the *mPing* insertion and the availability of *mPing* as an insertional mutagen in rice.

Materials and Methods

Plant materials

EG4 (cultivar Gimbozu) is a Japanese landrace temperate *japonica* cultivar exhibiting high *mPing* activity in nature (Naito et al. 2006). IM294 is a slender glume mutant line, which was induced by gamma irradiation of seeds of EG4, harboring a mutant allele *rurm1* at the *Rice ubiquitin-modifier 1* (*Rurm1*) locus destructed by an *mPing* insertion in exon 4 (Nakazaki et al. 2003; Tsukiyama et al. 2013). HS110 and HS169, like IM294, were gamma-ray induced mutant lines from EG4, exhibiting early- and late-heading (flowering), respectively. HS110 harbors a mutant allele *hd1* at the *Heading date 1* (*Hd1*) locus (Yano et al. 2000; Kikuchi et al. 2003), whereas HS169 harbors a mutant allele *ehd1* (=*ef1-h*) completely disrupted by an *mPing* insertion at the *Early heading 1* (*Ehd1*) locus (Nishida et al. 2002; Saito et al. 2009). All the plant materials were grown at an experimental paddy field at Kyoto University, Kyoto, Japan.

DNA and RNA extraction

A leaf blade was sampled from each of five plants per strain/line 30 days after sowing (DAS), and genomic DNA was extracted by cetryltrimethylammonium bromide (CTAB) method (Murray and Thompson 1980). For RNA extraction, a leaf blade was sampled from each of five plants per strain/line at 45 DAS (for *Hd1* and *Ehd1*) or 110 DAS (for *Rurm1*). Total RNA was extracted by

Quick Prep Total RNA extraction Kit (GE Healthcare, Little Chalfont, UK). DNA and RNA were quantified with a spectrophotometer (Biophotometer; Eppendorf, Hamburg, Germany), and stored at –20°C until use.

3’-RACE (Rapid Amplification of cDNA Ends) analysis

cDNA was synthesized in 20 μl reaction mixture containing 1 μg of total RNA, AMV Reverse Transcriptase XL (Takara Bio, Shiga, Japan), and oligo dT-3 site adaptor primer (Takara Bio). Synthesis conditions were as follows: 10 min at 30ºC, 30 min at 50ºC, 5 min at 95ºC, and 5 min at 5ºC. The 1st 3’-RACE reactions and the 2nd 3’-RACE reactions were performed with primers specific for each target gene and 3’ adaptor primer. The primer sequences and annealing temperatures for each primer are listed in ESM Table 1. Amplified fragments were subcloned into pGEM-T easy vector (Promega, Madison, WI, USA), and were sequenced using an ABI 3730xl DNA analyzer (Applied Biosystems, Foster City, CA, USA). Transcript isoforms were designated according to the nomenclature of McCouch (2008).

Sequence analyses

Sequences of the 3’-RACE products were analyzed using the ORF Finder program of the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). The

Plant *cis*-acting regulatory DNA elements (PLACE) database search (http://www.dna.affrc.go.jp/PLACE/) (Higo et al. 1999) was performed to identify polyadenylation signals in the *mPing* sequences. In rice, twenty hexamers (AATAAA, ATATAT, AAATAA, AATAAT, ATAAAA, TATATA, ATAAAT, TGAAAT, AATATA, ATGAAT, TAATAA, AATGAA, AATTTT, ATAATA, AAATTT, TTAATT, TTTGTT, AAAAAT, GAATAA, and AAATAT) have been reported as major polyadenylation signals (Shen et al. 2008). The twenty hexamers were also searched in the *mPing* sequences using the ClustalW program (http://clustalw.ddbj.nig.ac.jp/) (Thompson et al. 1994) of the DNA Data Bank of Japan (DDBJ). Functional domains and/or motifs in the *mPing* sequences were searched using the Pfam database (http://pfam.xfam.org/) (Finn et al. 2014).

Bisulfite sequencing

Genomic DNA was treated with sodium bisulfite using an EZ DNA Methylation Gold Kit (Zymo Research, Orange, CA, USA). Primers for bisulfite PCR (*hd1*-BS-F1: 5'-GAYAGTAAAAAAGATATTGGAAGTT-3' and *hd1*-BS-R1: 5'-CACCCTRRCCTCCCTRTCCAT-3') were designed with a Kismeth Primer Design program (http://katahdin.mssm.edu/kismeth/primer_design.pl) (Gruntman et al. 2008). Bisulfite PCR was performed in 50 μl reaction mixture containing 1 × EpiTaq PCR Buffer, 2.5 mM $MgCl_2$, 0.3 mM dNTP, 0.4 μM of each primer, 10 ng of bisulfite-treated DNA, and 1.25 U of TaKaRa EpiTaq HS

(Takara Bio). PCR conditions were as follows: 40 cycles of a denaturation step for 10 s at 95°C, an annealing step for 30 s at 50°C, and an extension step for 1 min at 72°C. PCR products were purified with Diffinity 2 (Sigma, USA) and cloned into pGEM-T vector (Promega). More than 10 clones were sequenced using an ABI 3730xl DNA analyzer (Applied Biosystems). Methylation degree was analyzed using a Kismeth Bisulfite Analysis program (http://katahdin.mssm.edu/kismeth/revpage.pl).

Results

Structure of the *hd1* transcripts in a mutant line HS110

Heading date 1 (*Hd1*) gene plays important roles in the causal genetic pathway of flowering (heading) in rice (Yano et al. 2000), consisting of two exons and a single intron. The early heading time mutant line HS110, which was induced with gamma ray irradiation of seeds of EG4, harbors a mutant allele *hd1* disrupted by an *mPing* insertion (Kikuchi et al. 2003) (Fig. 1). Since except for this *mPing* insertion, *hd1* has the same sequence as *Hd1*, and the insertion position of *mPing* is intron 1, the function of *Hd1* should be retained by correct RNA splicing in HS110. Nevertheless, HS110 flowers 14 days earlier than the original strain EG4 (Tanisaka et al. 1992, Yano et al. 2000). This indicates that the *mPing* insertion may affect the function of gene even if its insertion position is not exon. Yano et al. (2000) showed that HS110 yielded two transcripts whereas EG4 yielded a single

transcript. We performed 3'-RACE to determine the structure of *Hd1*/*hd1* transcripts in EG4 and HS110; consequently, we confirmed that the structure of *Hd1* transcript in EG4 was the same as that in the sequenced cultivar Nipponbare (ESM Fig. 1). In HS110, we obtained three different transcripts, and named them *hd1-s1*, *hd1-s2*, and *hd1-s3*, respectively (Fig. 1). *hd1-s1* was the normal transcript that was produced by correct RNA splicing. *hd1-s2* and *hd1-s3* were alternatively spliced isoforms of *hd1* gene: the former contained a 26-bp sequence (nucleotides 1557-1582) in the 3'-terminal part of intron and a 261-bp sequence (nucleotides 1583-1843) in the 5'-terminal part of *mPing*, and the latter consisted of exon 1, a 26-bp sequence (nucleotides 1557-1582) in the 3' terminal part of intron 1 and a 404-bp sequence (nucleotides 1538-1941) in the 5' terminal part of *mPing*. It is therefore considered that the *mPing* inserted within an intron can be incorporated as an alternative exon, and can induce an alternative 5' splice site and an alternative polyadenylation site within its sequence. Furthermore, it is indicated that *mPing* most likely influences the usage of 3' splice sites.

Structure of the transcripts of genes harboring the *mPing* insertion within an exon

In the *hd1* allele, we found that the *mPing* inserted within an intron induced alternative splicing and alternative polyadenylation. We examined whether the *mPing* inserted within an exon also alters the structure of transcripts. In the previous study, we documented that a slender glume mutant line

IM294 has an *mPing* insertion in exon 4 of *Rice ubiquitin-related modifier-1* (*Rurm1*) gene, which is responsible for the mutation of slender glume (Nakazaki et al. 2003) (Fig. 2). Using the same way as in the analysis of *Hd1/hd1* transcripts, we found that EG4 produced only a normal *Rurm1* transcript produced by correct RNA splicing, although the truncation of 3'UTR was observed in some transcripts (ESM Fig. 2). On the other hand, IM294 harboring a mutant allele *rurm1* yielded four different transcripts (*rurm1-s1~s4*) (Fig. 2). In *rurm1-s1*, three introns were correctly spliced out, and a whole *mPing* sequence was included in exon 4. In *rurm1-s2* and *rurm1-s3*, alternative polyadenylation occurred at different positions in the *mPing* sequence. *rurm1-s4* consisted of exon 1, exon 2, and a 50-bp sequence (nucleotides 636-685) in the 5'-terminal part of intron 3 that were retained by reading through the 5' splice site and by generating the alternative polyadenylation.

A late heading time mutant line HS169 has a mutant allele *ehd1-h* completely disrupted by an *mPing* insertion into exon 2 (Nishida et al. 2002; Saito et al. 2009) (Fig. 3). The *Ehd1* gene in the wild type consists of five exons and four introns. In EG4, we identified two alternatively spliced isoforms (*Ehd1-s2* and *Ehd1-s3*) along with the normal transcript (*Ehd1-s1*) (ESM Fig. 3). In *Ehd1-s2* and *Ehd1-s3*, intron 2 was retained by alternative 3' splicing, and alternative polyadenylation occurred at two different positions in intron 2. On the other hand, we obtained eight alternatively spliced transcripts (*ehd1-s1~s8*) from HS169 (Fig. 3). In *ehd1-s1*, a 160-bp sequence (nucleotides 1370-1529) in the 5'-terminal part of exon 2 was eliminated by being provided with

alternative 3’ splice site in the *mPing* sequence. In *ehd1-s2*, exon 2 having the *mPing* sequence was excluded from mature mRNA. In *ehd1-s3*, a 955-bp sequence (nucleotides 110-1064) in the 5’-terminal part of intron 1 was retained by reading through the 5’ splice site and by generating alternative polyadenylation. Furthermore, in *ehd1-s4~s8*, intron 1 was partially eliminated (nucleotides 110-693) by alternative 3’ splicing, and alternative polyadenylation occurred at different positions in alternatively retained intron 1. In addition to the results of 3’ RACE for the *rurm1* allele, these results indicate that the *mPing* inserted within an exon induces not only alternative 3’ splice and alternative polyadenylation sites within its sequence but also exon skipping. Furthermore, it is considered that *mPing* most likely influences the splicing pattern of intron adjacent to exon.

Premature termination codons and polyadenylation signals in the *mPing* sequence

Sequence analysis revealed that *mPing* has 23 and 16 potential premature termination codons (PTCs) (TAA, TAG, and TGA) on the plus and minus strands, respectively (Fig. 4 and ESM Fig. 4). The *Rurm1* gene encodes a 99 amino acid protein homologous to the yeast Urm1 (Ubiquitin-related modifier) protein (Furukawa et al. 2000). The C-terminal glycine-glycine residues are essential for the function of the Urm1 protein (Furukawa et al. 2000). Sequence analysis showed that the RURM1 proteins which were translated from *rurm1-s1*, *rurm1-s2*, and *rurm1-s3* lacked the C-terminal glycine residues due to a PTC within the *mPing* sequence (ESM Fig. 5). Furthermore, *ehd1-s1* also

had a PTC within the *mPing* sequence (ESM Fig. 6). To investigate whether the retained *mPing* sequence could provide the genes with a new functional activity, we searched functional domains and/or motifs on the *mPing* sequence by Pfam analysis. *mPing* encoded no domain and motif showing similarity to any known functional protein, indicating that proteins that were translated from mRNAs having the *mPing* sequence would not acquire any known functional activity. On the other hand, *hd1-s2* and *hd1-s3* were found to harbor a PTC within the alternatively retained intron sequence. The Hd1 protein has a CCT domain, which is often found near the C-terminus of proteins involved in photo-response signaling (Strayer et al. 2000). It was therefore considered that the Hd1 proteins translated from *hd1-s2* and *hd1-s3* might lose the function due to lacking the CCT domain (ESM Fig. 7).

Alternative polyadenylation was induced in the *mPing* sequences of *hd1* and *rurm1* alleles. In plants, polyadenylation is mainly regulated by polyadenylation signals, such as AATAAA and ATTATT, which are usually located 10- to 35-bp upstream of the cleavage site of 3' UTR (Wu et al. 1995; Shen et al. 2008). Using the PLACE database, we detected two (AATAAA and AATTAAA) and one (AATAAT) polyadenylation signals on the plus and minus strands of *mPing*, respectively (Fig. 4 and ESM Fig. 4). The locations of these signals, however, were far from the cleavage sites observed in *hd1-s3*, *rurm1-s2*, and *rurm1-s3*, respectively. In addition to AATAAA, 19 hexemers are known as major putative polyadenylation signals in rice (Shen et al. 2008). We searched these

hexamers in the 10- to 35-bp upstream regions of cleavage site in *hd1-s3*, *rurm1-s2*, and *rurm1-s3*. We detected one putative polyadenylation signal (ATAATA) in the 23-bp and 25-bp upstream regions of cleavage site in *hd1-s3* and *rurm1-s2*, respectively (Fig. 4). However, we detected no such hexamer in *rurm1-s3*. This indicates the existence of another polyadenylation signal in the *mPing* sequence.

DNA methylation of *mPing* and its flanking regions in the *hd1* allele

A recent study showed that DNA hypermethylation regulated the inclusion of alternative spliced exon (Maunakea et al. 2013). We conceived that *mPing* and/or its flanking regions might be hypermethylated in the mutant allele. To confirm this hypothesis, we investigated DNA methylation status of intron of the *Hd1/hd1* gene using bisulfite sequencing. All types of cytosine residues (CG, CHG, and CHH) were hardly methylated in the intron of the *Hd1* gene (Fig. 5), whereas in the *hd1* gene, the body region of *mPing* was heavily methylated at CG (98%) sites and moderately methylated at CHG (48%) and CHH (24%) sites. Moreover, the methylation level of the 5' *mPing*-flanking region highly increased in the *hd1* gene (Fig. 5). This region coincided with a part of the retained intron in *hd1-s2* and *hd1-s3*. This indicates that the alternative exon induced by *mPing* insertion might be regulated by DNA methylation targeting to *mPing* and/or its flanking regions.

Discussion

Nowadays, TEs are considered to contribute to the evolution of regulatory networks by altering gene expression at both the transcriptional and post-transcriptional levels (Feschotte 2008). In maize, TEs such as *Ds1* and *Mu1* are known to induce alternative splicing or alternative polyadenylation (Wessler et al. 1987; Ortiz and Strommer 1990; Wessler 1991). In *Solanacea*, the insertion of *MiS* element provides a functionally indispensable alternative exon in the tobacco mosaic virus N resistance gene (Kuang et al. 2009). These are experimental evidences that TEs have the capacity to alter regulatory networks at post-transcriptional level. Our previous studies showed that the rice active MITE *mPing* renders adjacent genes stress-inducible when it is inserted within 0.5-kb upstream of the transcription start site (Naito et al., 2009; Yasuda et al., 2013). In this study, we found that *mPing* induces alternative splicing and alternative polyadenylation when it is inserted within the coding sequence of genes. These findings demonstrate that *mPing* can alter gene expression not only at the transcriptional level but also at the post-transcriptional level.

HS66, like HS110, is an early heading mutant line, which was induced by gamma irradiation of seeds of EG4, and harbors a mutant allele *hd1* at the *Hd1* locus destructed by a 43-bp deletion in the first exon (Yano et al. 2000). Although HS66 produced the same amount of *hd1* transcript as EG4, the transcript had a PTC due to the 43-bp deletion (Yano et al. 2000). On the other hand, in HS110, small amount of functional transcript (*hd1-s1* in this study) was produced along

with various aberrant transcripts (probably, including *hd1-s2* and *hd1-s3* in this study) (Yano et al. 2000). Days to heading of HS110 was 4 days later than that of HS66 under natural field conditions (Yano et al. 2000). Yano et al. (2000) concluded that this phenotypic difference might reflect the presence of normal-size transcripts (*hd1-s1* in this study) in HS110. These findings support that the production of alternatively spliced transcripts due to the *mPing* insertion causes the disruption of the functional allele.

Alternative polyadenylation is recognized as a widespread mechanism of controlling gene expression, since the 3' UTR length influences the fate of mRNAs in several ways (Di Giammartino et al. 2011). In human, 6% of TEs (~1,500 TEs) give rise to polyadenylation sites (Chen et al. 2009). In rice, 280 genes used polyadenylation signals within MITE-derived sequences (Oki et al. 2008). Furthermore, in Arabidopsis, *COPIA-R7* inserted into the disease resistance gene *RPP7* affects the choice between two alternative polyadenylation sites in the *RPP7* pre-mRNA, and thereby influences the critical balance between RPP7-coding and non-RPP7-coding isoforms (Tsuchiya and Eulgem 2013). These findings indicate that TEs can drive the evolution of post-transcriptional regulation networks by providing genes with polyadenylation sites. In this study, it was shown that *mPing* could induce alternative polyadenylation sites within its own sequence. We investigated only the *mPing* inserted in the coding sequences of genes. In actuality, however, the EG4 genome includes 26 genes harboring the *mPing* insertion in the 3' UTR (Naito et al. 2009). The expression of

these genes is considered to be influenced by alternative polyadenylation signals provided by the *mPing* sequence. In this way, alternative polyadenylation induced by *mPing* also might play an important role in diversifying gene expression in rice.

In addition to polyadenylation signals, many potential PTCs are distributed on both plus and minus strands of *mPing*. Thus, *mPing* appears to easily produce transcripts encoding truncated proteins by providing PTC, independent of insertion direction, when the *mPing* sequences will be incorporated into mature mRNAs. In this study, *rurm1-s2* and *rurm1-s3* in IM294, and *ehd1-s1* in HS169 were expected to have PTC in the retained *mPing* sequences. Transcripts having PTC would be selectively degraded by the nonsense-mediated decay (NMD) pathway (Chang et al. 2007). However, the NMD pathway targets only transcripts having PTC at more than 55-bp upstream from the last exon/exon junction (Inacio et al. 2004; Hori et al. 2007). If the alternatively spliced transcripts have a PTC on the last exon, they could escape from the NMD pathway. In *rurm1-s2*, *rurm1-s3*, and *ehd1-s1*, *mPing* induced not only PTC but also alternative polyadenylation sites within its sequence, and consequently constituted the last exon having a PTC in mature transcript. Following the rules mentioned above, these transcripts would not be subject to degradation by the NMD pathway, and would produce truncated proteins. It was therefore considered that the exonization of *mPing* sequence could contribute to the proteome diversity, even if it leads to a truncated protein with loss-of-function or dominant-negative activities.

Recently, it has been reported that DNA methylation is involved in regulation of alternative splicing. In mammals, intragenic DNA methylation operates in exon definition to modulate alternative splicing and can enhance exon recognition via recruitment of a methyl-CpG binding protein MeCP2 (Maunakea et al. 2013). On the other hand, DNA methylation has been studied as the epigenetic defense mechanism of the host genome against active TEs since a long time ago. Like other TEs, *mPing* is known to be methylated in many cultivars (Shen et al. 2006; Ngezahayo et al. 2009; Wang et al. 2009). In this study, we found that, in the *hd1* gene, CG sites of *mPing* were heavily methylated, whereas CHG and CHH sites of 5' *mPing*-flanking region were moderately methylated compared with these of corresponding region in the wild type *Hd1* gene. In the rice genome, CHG and CHH methylation in gene body were retained at low level (approximately 6% and 1%, respectively) (Zemach et al. 2010). On the other hand, approximately 35% of CHG and 4% of CHH sites were methylated in 5' TE-flanking region (Zemach et al. 2010). These indicate that an increase of methylation in intron of the *hd1* gene was certainly caused by the *mPing* insertion. Although the relationship between methylation status of CHG and CHH sites and exon recognition is not fully understood, DNA methylation would be responsible for alternative splicing events in the *hd1* gene.

Alternative splicing is regulated by various abiotic stresses. Most of the genes that produce alternatively spliced transcripts in response to abiotic stresses are involved in the translational and

post-translational regulations (Mastrangelo et al. 2012). The *OsDREB2B* gene was found to produce two splice variants in response to drought and heat stresses in rice (Matsukura et al. 2010). In Arabidopsis, the loss-of-functions of *STA1* (Lee et al. 2006) and *RDM16* (Huang et al. 2013), both of which were pre-mRNA-splicing factors, caused hypersensitivity to cold and salt stresses, respectively. These indicate that alternative splicing is one of important mechanisms for plants to adapt to abiotic stress environments. In the RT-PCR assay for the *hd1* transcripts, HS110 produced different banding patterns in response to the transition from long-day to short-day conditions (Yano et al. 2000). This indicates that alternative splicing patterns of the *hd1* gene were altered by environmental condition. Although further experiments are needed, it is probable that *mPing* can change alternative splicing patterns in response to abiotic stresses.

In rice breeding, active TEs *Tos17* and *As/Ds* are employed for gene tagging systems because they disrupt gene functions by their transpositions into coding sequences (Miyao et al. 2003; Kolesnik et al. 2004). *mPing* can also induce by the transposition into coding sequences (Nakazaki et al. 2003; Kikuchi et al. 2003; Saito et al. 2009). However, unlike *Tos17* and *Ac/Ds*, *mPing* preferentially transposes into within 0.5-kb upstream of gene, and renders adjacent genes stress inducible (Naito et al. 2009; Yasuda et al. 2013). Although *mPing* is quiescent in most cultivars under natural growth conditions, the transposition of *mPing* can be transiently induced by various stress treatments (Kikuchi et al. 2003; Jiang et al. 2003; Nakazaki et al. 2003; Shan et al. 2005; Lin

et al. 2006). Furthermore, *mPing* is actively transposing without any stresses in several *japonica* landraces under natural growth condition, and its copy number reaches approximately 1000 copies (Naito et al. 2006). We have already established a screening system that detects *mPing* insertion near or in the target genes (Yasuda et al. 2013). In this study, we demonstrated that *mPing* induces alternative splicing and alternative polyadenylation, and thereby can influence gene expression at post-transcriptional level. Thus, we conclude that *mPing* would be a suitable element for mutagenesis in rice since it is able not only to produce loss-of-function alleles but also to modify the expression of a target gene at both the transcriptional and post-transcriptional levels.

Acknowledgement

This work was supported by grants from the Ministry of Education, Culture, Sports and Technology of Japan in the form of Grants-in-Aid for Scientific Research, 17380003 and 25292006.

Reference

Aguilar-Hernández V, Guzmán P (2013) Spliceosomal introns in the 5' untranslated region of plant BTL RING-H2 ubiquitin ligases are evolutionary conserved and required for gene expression. BMC Plant Biol 13:179

Ast G (2004) How did alternative splicing evolve? Nat Rev Genet 5: 773-782.

362 Cazzola M, Skoda RC (2000) Translational pathophysiology: a novel molecular mechanism of
363 human disease. Blood 95:3280-3288
364 Chan MT, Yu SM (1998) The 3' untranslated region of a rice alpha-amylase gene functions as a
365 sugar-dependent mRNA stability determinant. Proc Natl Acad Sci USA 95:6543-6547
366 Chang Y-F, Imam JS, Wilkinson MF (2007) The nonsense-mediated decay RNA surveillance
367 pathway. Annu Rev Biochem 76:51–74
368 Chen C, Ara T, Gautheret D (2009) Using *Alu* elements as polyadenylation sites: A case of
369 retroposon exaptation. Mol Biol Evol 26:327-334
370 Di Giammartino DC, Nishida K, Manley JL (2011) Mechanisms and consequences of alternative
371 polyadenylation. Mol Cell 2011 43:853-866
372 Feschotte C (2008) Transposable elements and the evolution of regulatory networks. Nat Rev Genet
373 9:397-405
374 Feschotte C, Pritham EJ (2007) DNA transposons and the evolution of eukaryotic genomes. Annu
375 Rev Genet 41:331-368
376 Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K,
377 Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M (2014) Pfam: the protein families database.
378 Nucleic Acids Res 42(Database issue):D222-230
379 Furukawa K, Mizushima N, Noda T, Ohsumi Y (2000) A protein conjugation system in yeast with

380 homology to biosynthetic enzyme reaction of prokaryotes. J Biol Chem 275:7462-7465

381 Gruntman E, Qi Y, Slotkin RK, Roeder T, Martienssen RA, Sachidanandam R (2008) Kismeth:

382 analyzer of plant methylation states through bisulfite sequencing. BMC Bioinformatics 9:371

383 Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant *cis*-acting regulatory DNA elements

384 (PLACE) database Nucleic Acids Res 27:297-300

385 Hori K, Watanabe Y (2007) Context Analysis of Termination Codons in mRNA that are Recognized

386 by Plant NMD. Plant Cell Physiol 48:1072–1078

387 Huang CF, Miki D, Tang K, Zhou HR, Zheng Z, Chen W, Ma ZY, Yang L, Zhang H, Liu R, He XJ,

388 Zhu JK (2013) A Pre-mRNA-splicing factor is required for RNA-directed DNA methylation in

389 *Arabidopsis*. PLoS Genet 9:e1003779

390 Inacio A, Silva AL, Pinto J, Ji X, Morgado A, Almeida F, Faustino P, Lavinha J, Liebhaber SA,

391 Romao L (2004) Nonsense mutations in close proximity to the initiation codon fail to trigger full

392 nonsense-mediated mRNA decay. J. Biol. Chem 279:32170–32180

393 Jiang N, Bao Z, Zhang X, Hirochika H, Eddy SR, McCouch SR, Wessler SR (2003) An active DNA

394 transposon family in rice. Nature 421:163-167

395 Kikuchi K, Terauchi K, Wada M, Hirano HY (2003) The plant MITE *mPing* is mobilized in anther

396 culture. Nature 421:167-170

397 Kolesnik T, Szeverenyi I, Bachmann D, Kumar CS, Jiang S, Ramamoorthy R, Cai M, Ma ZG,

398 Sundaresan V, Ramachandran S (2004) Establishing an efficient *Ac/Ds* tagging system in rice:
399 large-scale analysis of *Ds* flanking sequences. Plant J 37:301–314
400 Kornblihtt AR, Schor IE, Alló M, Dujardin G, Petrillo E, Muñoz MJ (2013) Alternative splicing: a
401 pivotal step between eukaryotic transcription and translation. Nat Rev Mol Cell Biol 14:153-165
402 Kuang H, Padmanabhan C, Li F, Kamei A, Bhaskar PB, Ouyang S, Jiang J, Buell CR, Baker B
403 (2009) Identification of miniature inverted-repeat transposable elements (MITEs) and biogenesis
404 of their siRNAs in the *Solanaceae*: new functional implications for MITEs. Genome Res
405 19:42-56
406 Lee BH, Kapoor A, Zhu J, Zhu JK (2006) STABILIZED1, a stress-upregulated nuclear protein, is
407 required for pre-mRNA splicing, mRNA turnover, and stress tolerance in *Arabidopsis*. Plant Cell
408 18:1736-1749
409 Lin X, Long L, Shan X, Zhang S, Shen S, Liu B (2006) In planta mobilization of *mPing* and its
410 putative autonomous element *Pong* in rice by hydrostatic pressurization. J Exp Bot 57:2313-2323
411 Lytle JR, Yario TA, Steitz JA (2007) Target mRNAs are repressed as efficiently by
412 microRNA-binding sites in the 5' UTR as in the 3' UTR. Proc Natl Acad Sci USA 104:9667-9672
413 Mangone M, Manoharan AP, Thierry-Mieg D et al (2010) The landscape of *C. elegans* 3'UTRs.
414 Science 329:432-435
415 Marquez Y, Brown JW, Simpson C, Barta A, Kalyna M (2012) Transcriptome survey reveals

416 increased complexity of the alternative splicing landscape in *Arabidopsis*. Genome Res
417 22:1184-1195
418 Mastrangelo AM, Marone D, Laidò G, De Leonardis AM, De Vita P (2012) Alternative splicing:
419 Enhancing ability to cope with stress via transcriptome Plant Sci 185-186:40-49
420 plasticity.
421 Matsukura S, Mizoi J, Yoshida T, Todaka D, Ito Y, Maruyama K, Shinozaki K, Yamaguchi-Shinozaki
422 K (2010) Comprehensive analysis of rice DREB2-type genes that encode transcription factors
423 involved in the expression of abiotic stress-responsive genes. Mol Genet Genomics 283:185-196
424 Mazumder B, Sampath P, Fox PL (2005) Regulation of macrophage ceruloplasmin gene expression:
425 one paradigm of 3'-UTR-mediated translational control. Mol Cells 20:167-172
426 Maunakea AK, Chepelev I, Cui K, Zhao K (2013) Intragenic DNA methylation modulates
427 alternative splicing by recruiting MeCP2 to promote exon recognition. Cell Res 23:1256-1269
428 Mayr C, Bartel DP (2009) Widespread shortening of 3'UTRs by alternative cleavage and
429 polyadenylation activates oncogenes in cancer cells. Cell 138:673-684
430 McClintock B (1950) The origin and behavior of mutable loci in maize. Proc Natl Acad Sci USA
431 36:344-355
432 McCouch SR (2008) Gene nomenclature system for rice. Rice 1:72-84
433 Misquitta CM, Chen T, Grover AK (2006) Control of protein expression through mRNA stability in

calcium signalling. Cell Calcium 40:329-346

Miyao A, Tanaka K, Murata K, Sawaki H, Takeda S, Abe K, Shinozuka Y, Onosato K, Hirochika H (2003) Target site specificity of the *Tos17* retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. Plant Cell 15:1771–1780

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321-4325

Naito K, Cho E, Yang G, Campbell MA, Yano K, Okumoto Y, Tanisaka T, Wessler SR (2006) Dramatic amplification of a rice transposable element during recent domestication. Proc Natl Acad Sci USA 103:17620-17625

Naito K, Zhang F, Tsukiyama T, Saito H, Hancock CN, Richardson AO, Okumoto Y, Tanisaka T, Wessler SR (2009) Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. Nature 461:1130-1134

Nakazaki T, Okumoto Y, Horibata A, Yamahira S, Teraishi M, Nishida H, Inoue H, Tanisaka T (2003) Mobilization of a transposon in the rice genome. Nature 421:170-172

Ngezahayo F, Xu C, Wang H, Jiang L, Pang J, Liu B (2009) Tissue culture-induced transpositional activity of mPing is correlated with cytosine methylation in rice. BMC Plant Biol 9:91

Nilsen TW, Graveley BR (2010) Expansion of the eukaryotic proteome by alternative splicing.

452 Nature 463:457-463

453 Nishida H, Inoue H, Okumoto Y, Tanisaka T (2002) A novel gene *ef1-h* conferring an extremely long
454 basic vegetative growth period in rice. Crop Sci 42:348–354

455 Oki N, Yano K, Okumoto Y, Tsukiyama T, Teraishi M, Tanisaka T (2008) A genome-wide view of
456 miniature inverted-repeat transposable elements (MITEs) in rice, *Oryza sativa* ssp. *japonica*.
457 Genes Genet Syst 83:321-329

458 Ortiz DF, Strommer JN (1990) The *Mu1* maize transposable element induces tissue-specific aberrant
459 splicing and polyadenylation in two *Adh1* mutants. Mol Cell Biol 10:2090-2095

460 Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ (2008) Deep surveying of alternative splicing
461 complexity in the human transcriptome by high-throughput sequencing. Nat Genet 40:1413-1415

462 Saito H, Yuan Q, Okumoto Y, Doi K, Yoshimura A, Inoue H, Teraishi M, Tsukiyama T, Tanisaka T
463 (2009) Multiple alleles at *Early flowering 1* locus making variation in the basic vegetative growth
464 period in rice (*Oryza sativa* L.). Theor Appl Genet 119:315-323

465 Schnable P, Ware D, Fulton R et al (2009) The B73 maize genome: complexity, diversity, and
466 dynamics. Science 326:1112–1115

467 Shan X, Liu Z, Dong Z, Wang Y, Chen Y, Lin X, Long L, Han F, Dong Y, Liu B (2005) Mobilization
468 of the active MITE transposons *mPing* and *Pong* in rice by introgression from wild rice (*Zizania*
469 *latifolia* Griseb.). Mol Biol Evol 22:976-990

Shen S, Wang Z, Shan X, Wang H, Li L, Lin X, Long L, Weng K, Liu B, Zou G (2006) Alterations in DNA methylation and genome structure in two rice mutant lines induced by high pressure. Sci China C Life Sci 49:97-104

Shen Y, Ji G, Haas BJ, Wu X, Zheng J, Reese GJ, Li QQ (2008) Genome level analysis of rice mRNA 3'-end processing signals and alternative polyadenylation. Nucleic Acid Res 36:3150-3161

Strayer C, Oyama T, Schultz TF, Raman R, Somers DE, Más P, Panda S, Kreps JA, Kay SA (2000) Cloning of the *Arabidopsis* clock gene *TOC1*, an autoregulatory response regulator homolog. Science 289:768-771

Syed NH, Kalyna M, Marquez Y, Barta A, Brown JW (2012) Alternative splicing in plants--coming of age. Trends Plant Sci 17:616-623

Tanisaka T, Inoue H, Uozu S, Yamagata H (1992) Basic vegetative growth and photoperiod sensitivity of heading-time mutants induced in rice. Japan J Breed 42: 657-668.

Teramoto S, Tsukiyama T, Okumoto Y, Tanisaka T (2014) Early embryogenesis-specific expression of the rice transposon *Ping* enhances amplification of the MITE *mPing*. PLoS Genet 10(6):e1004396

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap

488 penalties and weight matrix choice. Nucleic Acids Res 22:4673-4680

489 Tsuchiya T and Eulgem T (2013) An alternative polyadenylation mechanism coopted to the

490 *Arabidopsis RPP7* gene through intronic retrotransposon domestication. Proc Natl Acad Sci USA

491 110:E3535-3543

492 Tsukiyama T, Teramoto S, Yasuda K, Horibata A, Mori N, Okumoto Y, Teraishi M, Saito H, Onishi A,

493 Tamura K, Tanisaka T (2013) Loss-of-function of a ubiquitin-related modifier promotes the

494 mobilization of the active MITE *mPing*. Mol Plant 6:790-801

495 Turcotte K, Srinivasan S, Bureau T (2001) Survey of transposable elements from rice genomic

496 sequences. Plant J 25:169-179

497 Wang H, Chai Y, Chu X, Zhao Y, Wu Y, Zhao J, Ngezahayo F, Xu C, Liu B (2009) Molecular

498 characterization of a rice mutator-phenotype derived from an incompatible cross-pollination

499 reveals transgenerational mobilization of multiple transposable elements and extensive epigenetic

500 instability. BMC Plant Biol 9:63

501 Wessler SR (1991) The maize transposable *Ds1* element is alternatively spliced from exon sequences.

502 Mol Cell Biol 11:6192-6196

503 Wessler SR, Baran G, Varagona M (1987) The maize transposable element *Ds* is spliced from RNA.

504 Science 237:916-918

505 Wu L, Ueda T, Messing J (1995) The formation of mRNA 3'-ends in plants. Plant J 8:323-329

506 Xing D, Li QQ (2011) Alternative polyadenylation and gene expression regulation in plants. Wiley
507 Interdiscip Rev RNA 2:445-458
508 Wu X, Liu M, Downie B, Liang C, Ji G, Li QQ, Hunt AG (2011) Genome-wide landscape of
509 polyadenylation in *Arabidopsis* provides evidence for extensive alternative polyadenylation. Proc
510 Natl Acad Sci USA 108:12533–12538
511 Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara
512 Y, Nagamura Y, Sasaki T (2000) *Hd1*, a major photoperiod sensitivity quantitative trait locus in
513 rice, is closely related to the *Arabidopsis* flowering time gene *CONSTANS*. Plant Cell
514 12:2473-2484
515 Yasuda K, Ito M, Sugita T, Tsukiyama T, Saito H, Naito K, Teraishi M, Tanisaka T, Okumoto Y
516 (2013) Utilization of transposable element *mPing* as a novel genetic tool for modification of the
517 stress response in rice. Mol Breed 32:505-516
518 Zemach A, Kim MY, Silva P, Rodrigues JA, Dotson B, Brooks MD, Zilberman D (2010) Local DNA
519 hypomethylation activates genes in rice endosperm. Proc Natl Acad Sci USA 107:18729-18734
520

Figure legends

Fig. 1 Schematic representation of *hd1* allele and its transcripts.

White, black, and gray boxes indicate exon, *mPing*, and retained intron, respectively. *hd1* allele is composed of two exons (nucleotides 1-987, 2058-2627), one intron (nucleotides 988-2057) and *mPing* (nucleotides 1583-2012) inserted in the intron. Horizontal line and black inverted triangle indicate intron and PTC, respectively. *hd1-s2* was generated by alternative 3' splicing of intron and alternative 5' splicing of exon 2 having the *mPing* sequence. *hd1-s3* was generated by alternative 3' splicing of intron and alternative polyadenylation.

Fig. 2 Schematic representation of *rurm1* allele and its transcripts.

White, black, and gray boxes indicate exon, *mPing*, and retained intron, respectively. Horizontal lines and black inverted triangle indicates intron and PTC, respectively. *rurm1* allele is composed of four exons (nucleotides 1-240, 352-430, 553-635, 1317-2077), three introns (nucleotides 241-351, 431-552, 634-1316) and *mPing* (nucleotides 1343-1772) inserted in the 4th exon. *rurm1-s2* and *-s3* are generated by alternative polyadenylation occurred at different position within *mPing* sequence. *rurm1-s4* has retained intron generated by alternative 3' splicing of intron 3.

Fig. 3 Schematic representation of *ehd1* allele and its transcripts.

White, black, and gray boxes indicate exon, *mPing*, and retained intron, respectively. *ehd1* allele is composed of five exons (nucleotides 1-109, 1370-1955, 3797-4198, 4348-4424, 5396-5680), four introns (nucleotides 110-1369, 1956-3796, 4199-4347, 4425-5395) and *mPing* (1383-1812) inserted in the 2nd exon. Horizontal lines and black inverted triangle indicates intron and PTC, respectively. *ehd1-s1* is generated by alternative 3' splicing of intron 1. *ehd1-s2* is generated by exon skipping of exon 2. *ehd1-s3* has retained intron generated by reading through the 5' splicing site of intron 1. Dark gray box in *ehd1-s3* indicates sequences spliced out in *ehd1-s4~s8*.

Fig. 4 PTCs and polyadenylation signals on the plus strand of *mPing*.

Black, gray and light gray boxes indicate TAA, TGA, and TAG, respectively. Polyadenylation signals detected by using PLACE database are underlined with dotted lines. Polyadenylation signals reported by Shen et al. (2008) are underlined with bold lines.

Fig. 5 Cytosine methylation of the *Hd1/hd1* locus in EG4 and HS110

a Dot plots of cytosine methylation in the *mPing*-body region and the *mPing*-flanking regions of the *Hd1/hd1* locus. Red, blue, and green circles indicate cytosine in the CG, CHG, and CHH sites, respectively. Filled and empty circles indicate methylated and ummethylated cytosines, respectively.

b, c, d Comparison of cytosine methylation degree between EG4 and HS110. Red, blue, and green

columns in the histograms represent the collective methylation degree of CG, CHG, and CHH sites, respectively, at the 5' *mPing*-flanking region (b), the *mPing*-body region (c), and the 3' *mPing*-flanking region (d).

EMS Fig. 1 Schematic representation of the *Hd1* allele and its transcripts.

White boxes and horizontal line indicate exon and intron, respectively. *Hd1-s1* is a normal transcript.

EMS Fig. 2 Schematic representation of the *Rurm1* allele and its transcripts.

White boxes and horizontal lines indicate exon and intron, respectively. *Rurm1*-s1 is a normal transcript.

EMS Fig. 3 Schematic representation of the *Ehd1* allele and its transcripts.

White and gray boxes indicate exon and retained intron, respectively. Horizontal lines and black inverted triangle indicate intron and PTC, respectively. *Ehd1-s1* is a normal transcript. *Ehd1-s2* and *Ehd1-s3* are alternative spliced isoforms generated by generated by reading through the 5' splicing site of intron 2.

EMS Fig. 4 PTCs and polyadenylation signals on the minus strand of *mPing*.

Black, gray and light boxes indicate TAA, TGA, and TAG, respectively. Polyadenylation signals detected by using PLACE database are underlined with a dotted line. Polyadenylation signals reported by Shen et al. (2008) are underlined with bold lines.

EMS Fig. 5 Deduced amino acid sequences of proteins translated from transcript isoforms of the *Hd1/hd1* allele.

Amino acid sequences of proteins translated from transcript isoforms of the *Hd1/hd1* allele are deduced by ORF finder. Asterisks represent stop codon.

EMS Fig. 6 Deduced amino acid sequences of proteins translated from transcript isoforms of the *Rurm1/rurm1* allele.

Amino acid sequences of proteins translated from transcript isoforms of the *Rurm1/rurm1* allele are deduced by ORF finder. Gray boxes indicate glycine residues which are necessary for the function. Asterisks represent stop codon.

EMS Fig. 7 Deduced amino acid sequences of proteins translated from transcript isoforms of the *Ehd1/ehd1* allele.

593 Amino acid sequences of proteins translated from transcript isoforms of the *Ehd1e/hd1* allele are

594 deduced by ORF finder. Asterisks represent stop codon.

595

596

ESM Table 1 Primer list for 3’-RACE

Target gene		Sequence	Annealing temperature (°C)
Hd1/hd1	1st	CGACAACCGCATCGAAAAC	60
	2nd	GAACAGCAAGAGCAGCAG	54
Ehd1/ehd1	1st	GGCCTTATGGACTAAGAGTTCTGG	58
	2nd	GACGACTGTTCATACTTGTCAGTCA	58
Rurm1/rurm1	1st	CACCATGCATCTAACCCTCGAATTCG	54
	2nd	GTCGTGATGAAAGGGTTGCTCG	54

Fig. 1

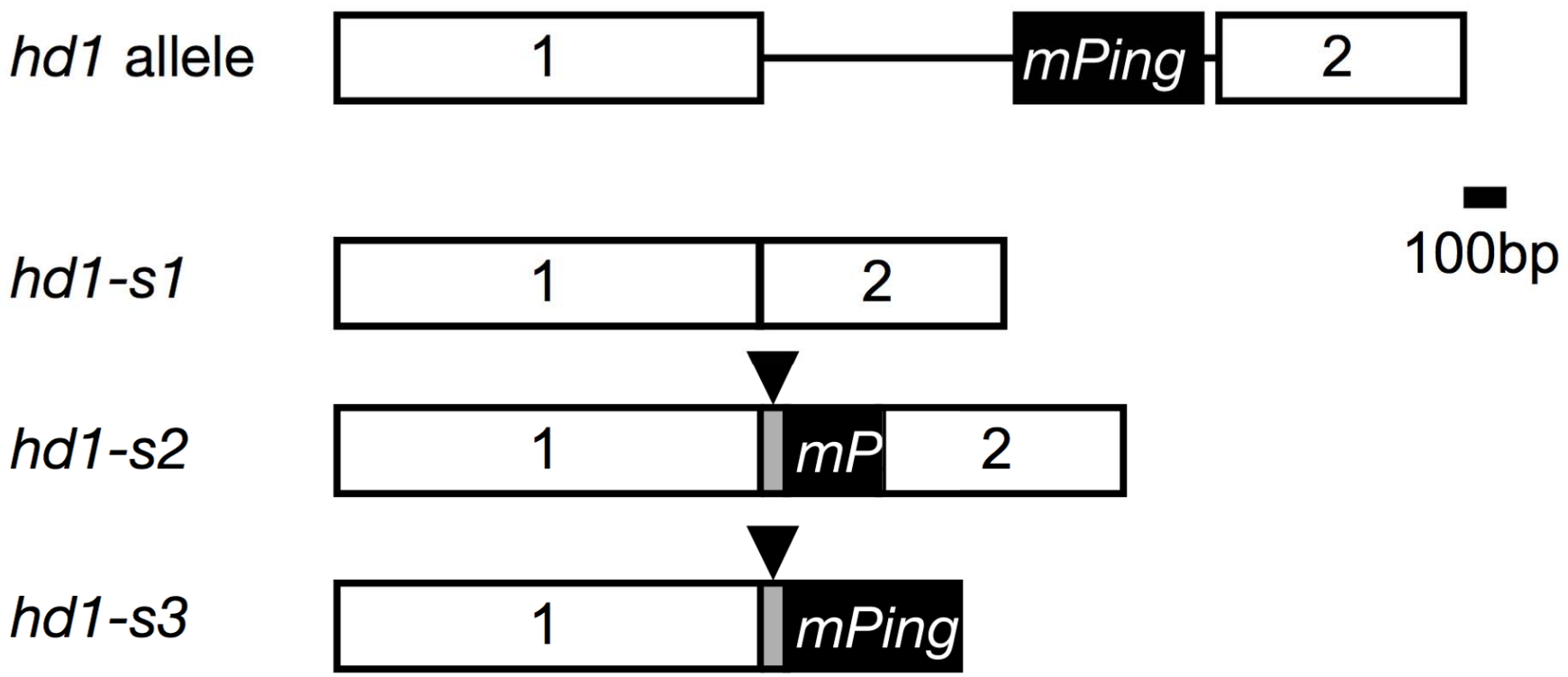


Fig. 2

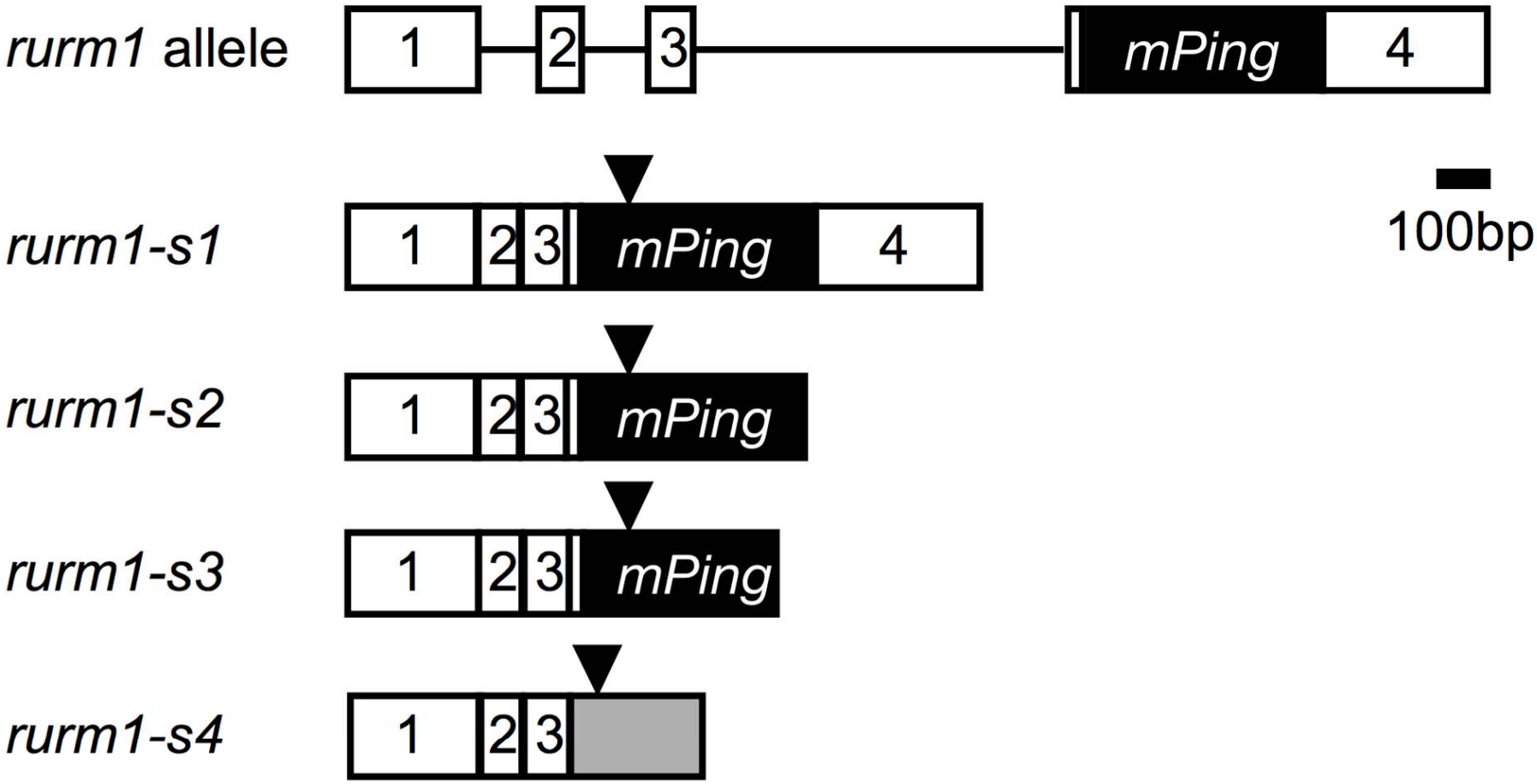

rurm1 allele
1
2
3
mPing
4
rurm1-s1
1
2
3
mPing
4
100bp
rurm1-s2
1
2
3
mPing
rurm1-s3
1
2
3
mPing
rurm1-s4
1
2
3

Fig. 3

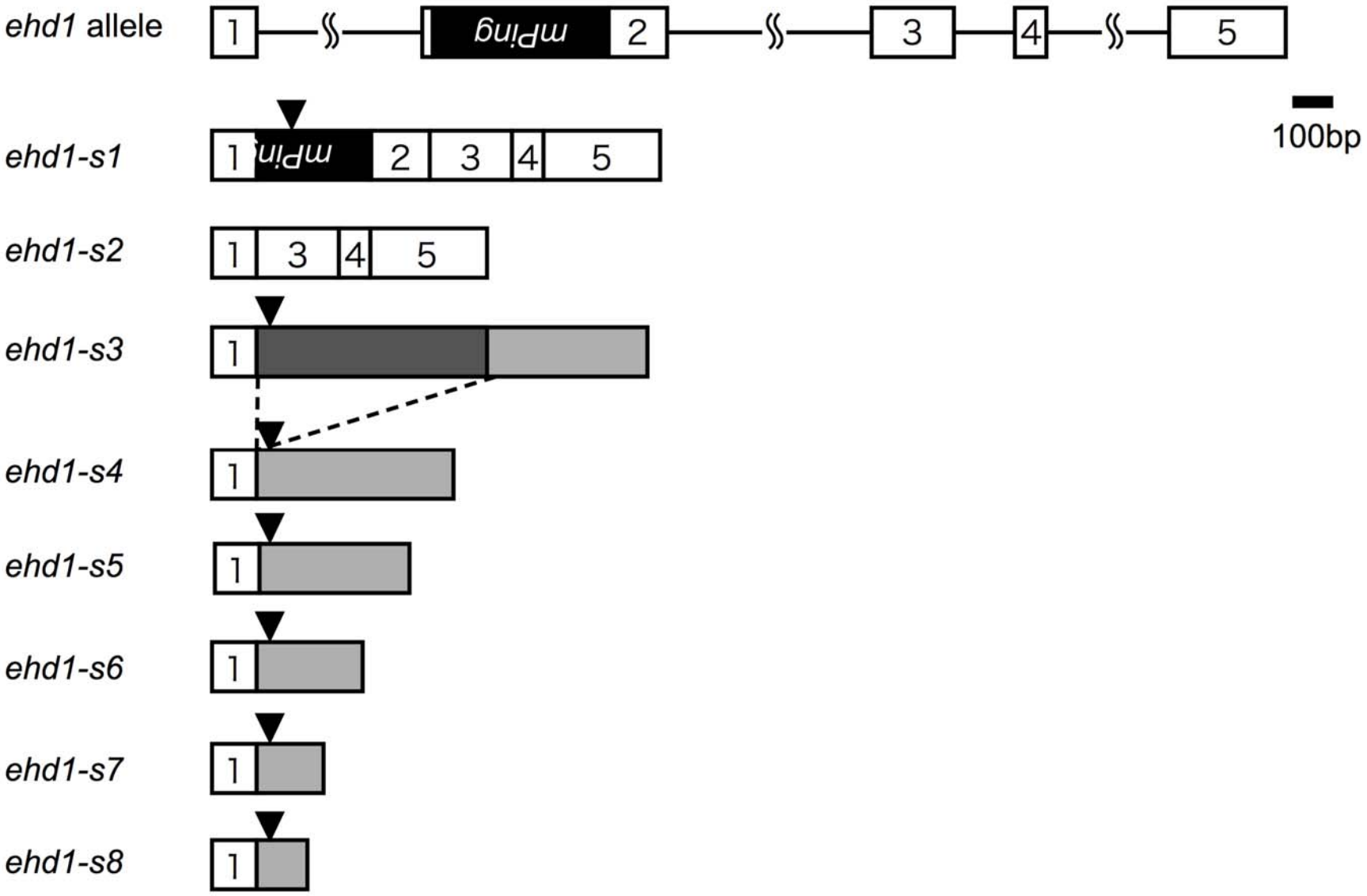

ehd1 allele
1
mPing
2
3
4
5
100bp
ehd1-s1
ehd1-s2
ehd1-s3
ehd1-s4
ehd1-s5
ehd1-s6
ehd1-s7
ehd1-s8

Fig. 4

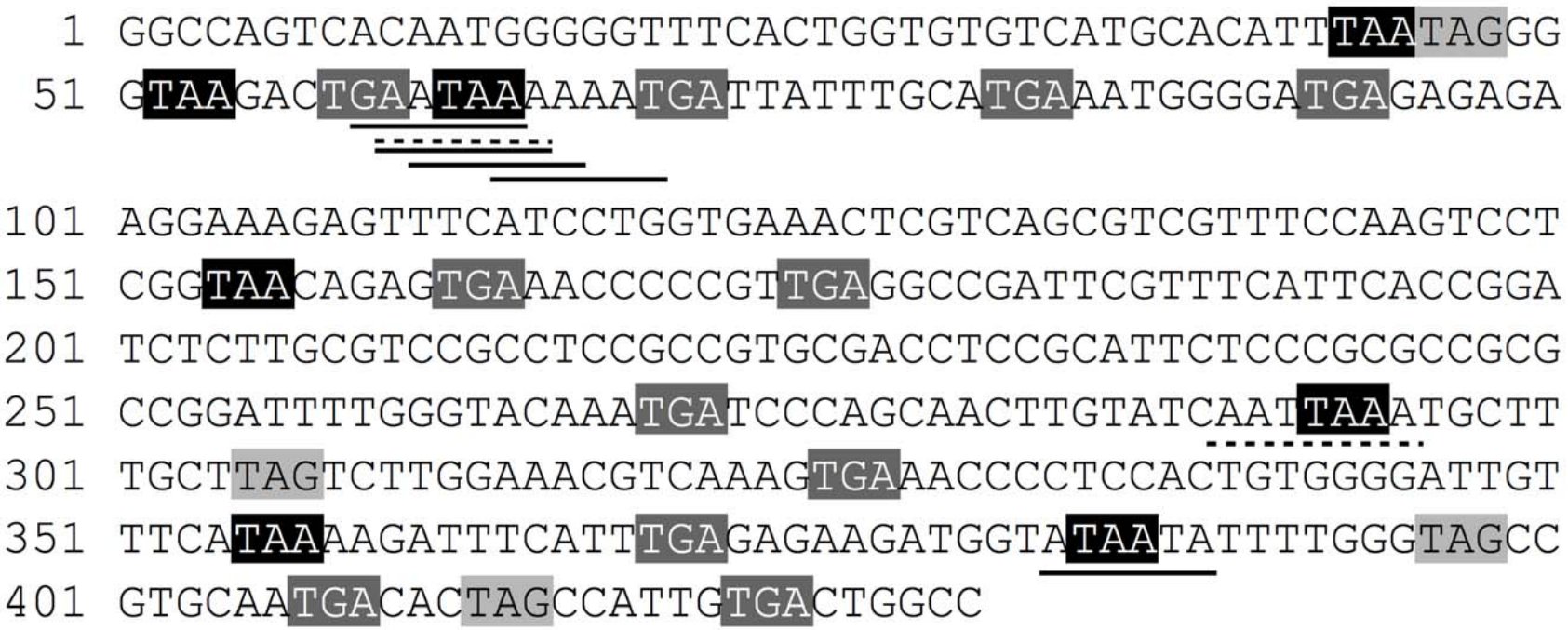
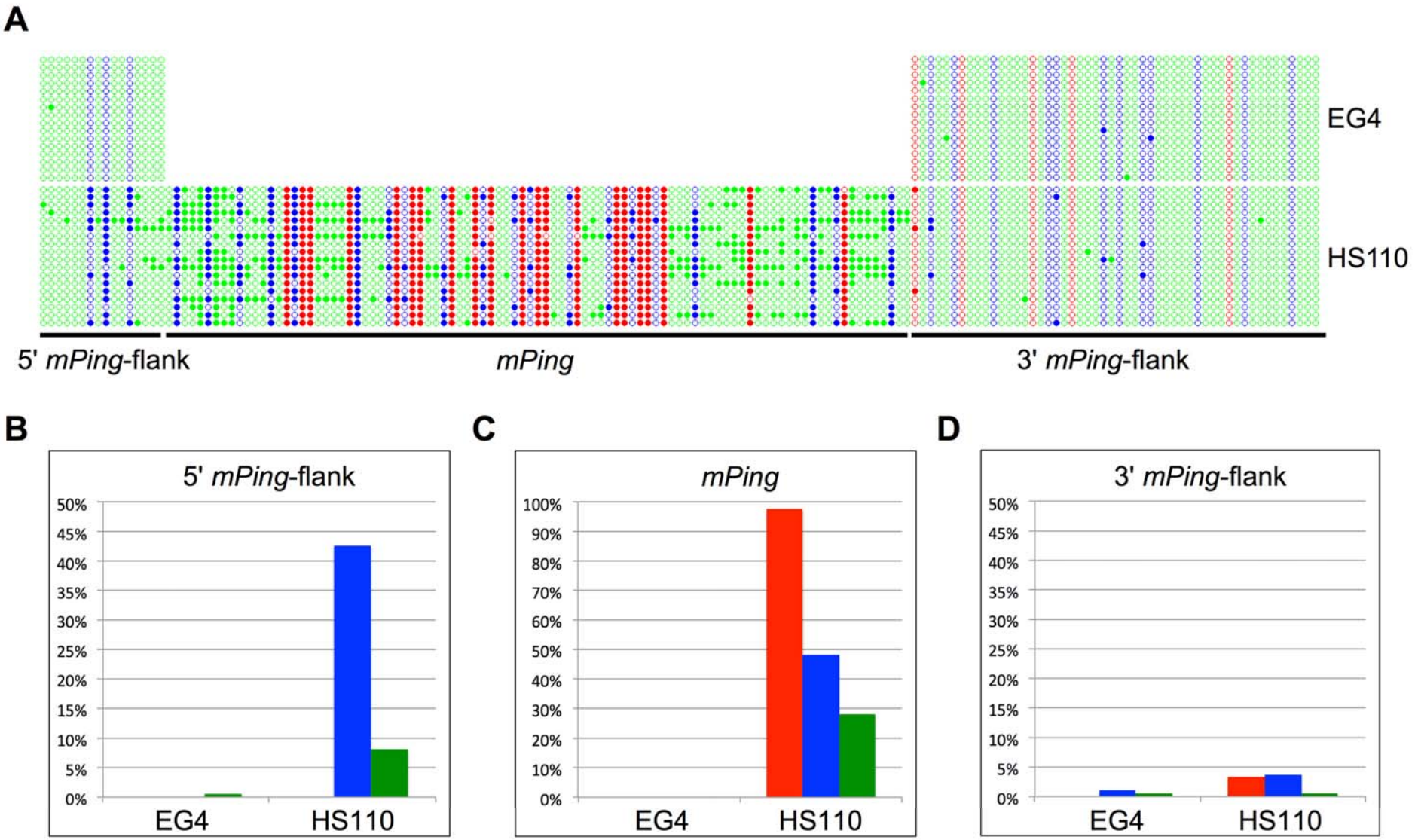
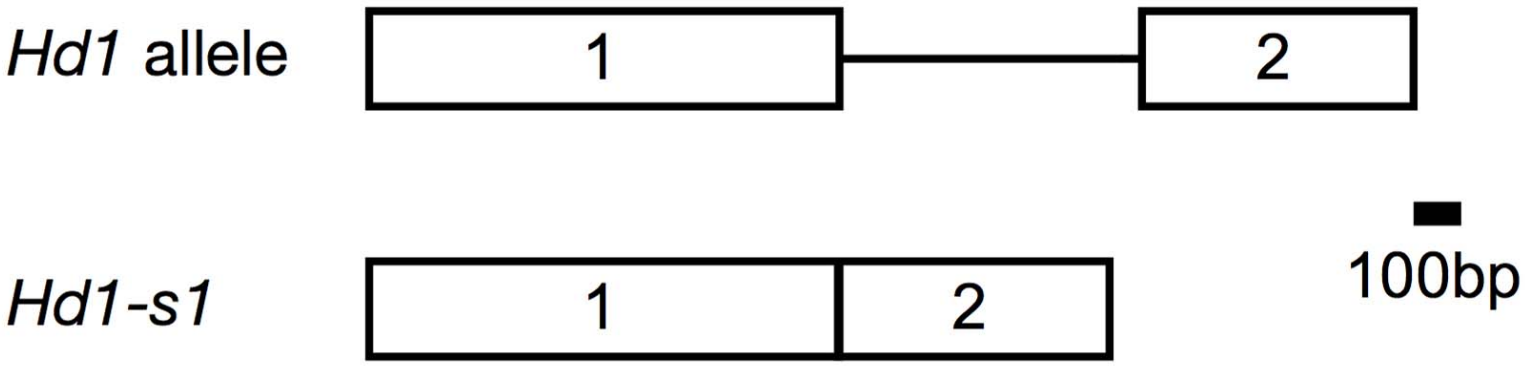

1 GGCCAGTCACAATGGGGGTTTCACTGGTGTGTCATGCACATTTAATAGGG
51 GTAAGACTGAATAAAAAATGATTATTTGCATGAAATGGGGATGAGAGAGA
101 AGGAAAGAGTTTCATCCTGGTGAAACTCGTCAGCGTCGTTTCCAAGTCCT
151 CGGTAACAGAGTGAAACCCCCGTTGAGGCCGATTCGTTTCATTCACCGGA
201 TCTCTTGCGTCCGCCTCCGCCGTGCGACCTCCGCATTCTCCCGCGCCGCG
251 CCGGATTTTGGGTACAAATGATCCCAGCAACTTGTATCAATTAAATGCTT
301 TGCTTAGTCTTGGAAACGTCAAAGTGAAACCCCTCCACTGTGGGGATTGT
351 TTCATAAAAGATTTCATTTGAGAGAAGATGGTATAATATTTTGGGTAGCC
401 GTGCAATGACACTAGCCATTGTGACTGGCC

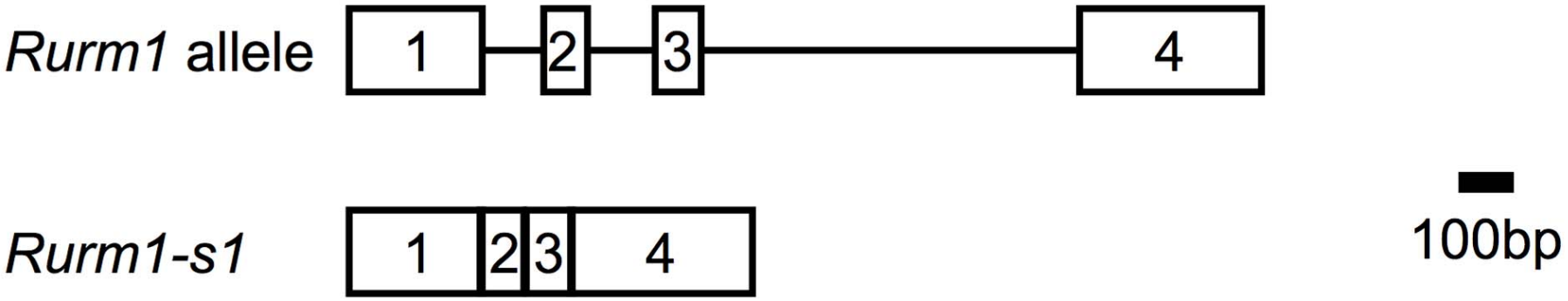
Fig. 5


A
EG4
HS110
5' mPing-flank
mPing
3' mPing-flank
B
5' mPing-flank
50%
45%
40%
35%
30%
25%
20%
15%
10%
5%
0%
EG4
HS110
C
mPing
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
EG4
HS110
D
3' mPing-flank
50%
45%
40%
35%
30%
25%
20%
15%
10%
5%
0%
EG4
HS110

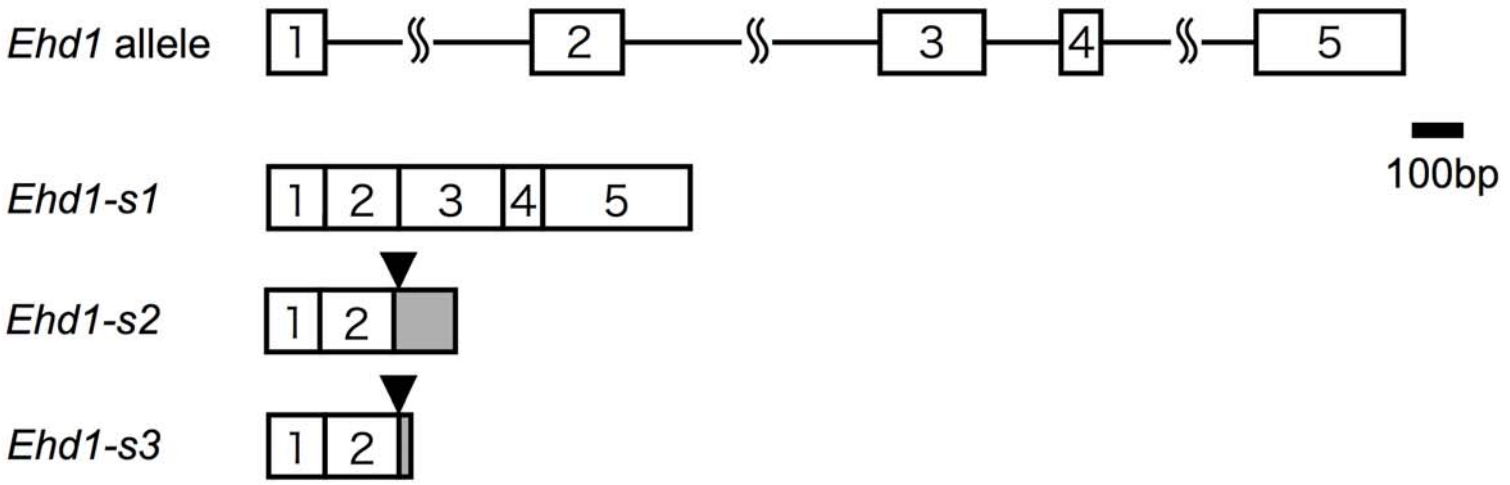
EMS_Fig. 1



EMS_Fig. 2



EMS_Fig. 3

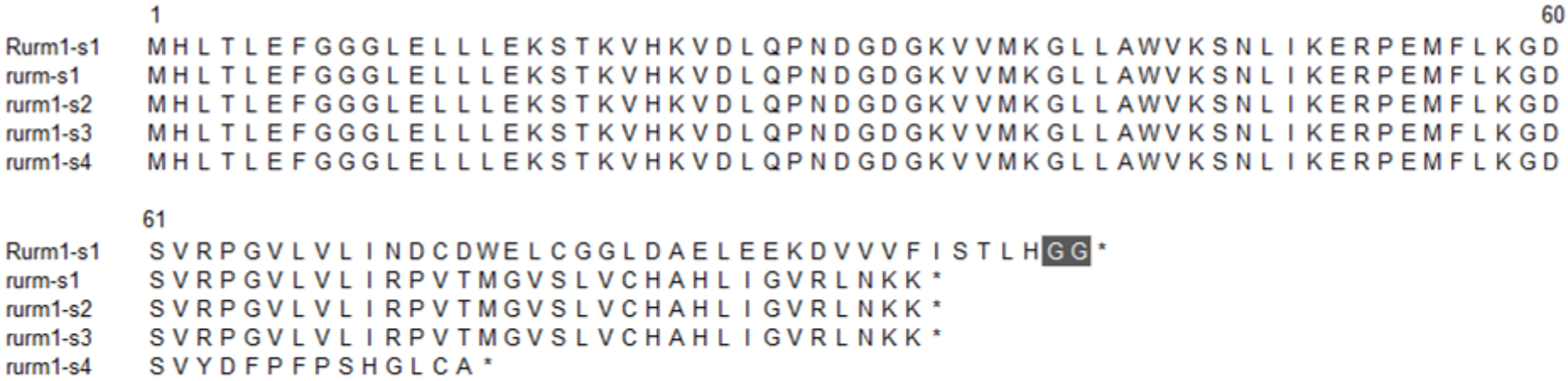


EMS_Fig. 4

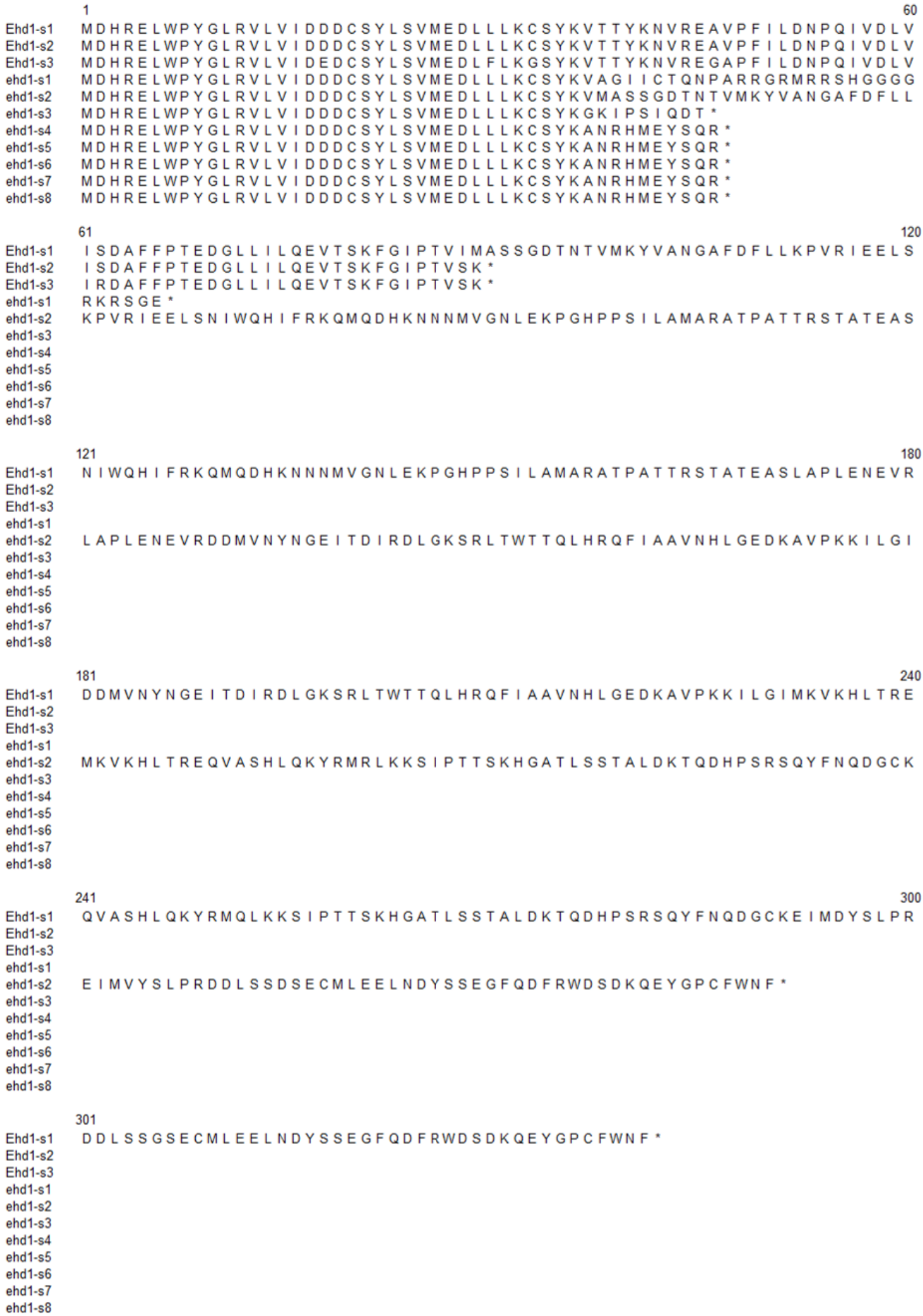

1 GGCCAGTCACAATGGCTAGTGTCATTGCACGGCTACCCAAAATATTATAC
51 CATCTTCTCTCAAATGAAATCTTTTATGAAACAATCCCCACAGTGGAGGG
101 GTTTCACTTTGACGTTTCCAAGACTAAGCAAAGCATTTAATTGATACAAG
151 TTGCTGGGATCATTTGTACCCAAAATCCGGCGCGGCGCGGGAGAATGCGG
201 AGGTCGCACGGCGGAGGCGGACGCAAGAGATCCGGTGAATGAAACGAATC
251 GGCCTCAACGGGGGTTTCACTCTGTTACCGAGGACTTGGAAACGACGCTG
301 ACGAGTTTCACCAGGATGAAACTCTTTCCTTCTCTCTCATCCCCATTTCA
351 TGCAAATAATCATTTTTTATTCAGTCTTACCCCTATTAAATGTGCATGAC
401 ACACCAGTGAAACCCCCATTGTGACTGGCC


1 60
Hd1-s1 MNYNFGGNVFDQEVGVGGEGGGGGEGSGCPWARPCDGCRAAPSVVYCRADAAYLCASCDA
hd1-s1 MNYNFGGNVFDQEVGVGGEGGGGGEGSGCPWARPCDGCRAAPSVVYCRADAAYLCASCDA
hd1-s2 MNYNFGGNVFDQEVGVGGEGGGGGEGSGCPWARPCDGCRAAPSVVYCRADAAYLCASCDA
hd1-s3 MNYNFGGNVFDQEVGVGGEGGGGGEGSGCPWARPCDGCRAAPSVVYCRADAAYLCASCDA
61 120
Hd1-s1 RVHAANRVASRHERVRVCEACERAPAALACRADAAALCVACDVQVHSANPLPAITIPATS
hd1-s1 RVHAANRVASRHERVRVCEACERAPAALACRADAAALCVACDVQVHSANPLPAITIPATS
hd1-s2 RVHAANRVASRHERVRVCEACERAPAALACRADAAALCVACDVQVHSANPLPAITIPATS
hd1-s3 RVHAANRVASRHERVRVCEACERAPAALACRADAAALCVACDVQVHSANPLPAITIPATS
121 180
Hd1-s1 VLAEAVVATATVLGDKDEEVDSWLLLSKDSDNNNNNNNNNNDNDNNDNNNSNSSNNGMYFG
hd1-s1 VLAEAVVATATVLGDKDEEVDSWLLLSKDSDNNNNNNNNNNDNDNNDNNNSNSSNNGMYFG
hd1-s2 VLAEAVVATATVLGDKDEEVDSWLLLSKDSDNNNNNNNNNNDNDNNDNNNSNSSNNGMYFG
hd1-s3 VLAEAVVATATVLGDKDEEVDSWLLLSKDSDNNNNNNNNNNDNDNNDNNNSNSSNNGMYFG
181 240
Hd1-s1 EVDEYFDLVGYNSYYDNRIENNQDRQYGMHEQQEQQQQQQEMQKEFAEKEGSECVVPSQI
hd1-s1 EVDEYFDLVGYNSYYDNRIENNQDRQYGMHEQQEQQQQQQEMQKEFAEKEGSECVVPSQI
hd1-s2 EVDEYFDLVGYNSYYDNRIENNQDRQYGMHEQQEQQQQQQEMQKEFAEKEGSECVVPSQI
hd1-s3 EVDEYFDLVGYNSYYDNRIENNQDRQYGMHEQQEQQQQQQEMQKEFAEKEGSECVVPSQI
241 300
Hd1-s1 TMLSEQQHSGYGVVGADQAASMTAGVSAYTDSISNSISFSSMEAGIVPDSTVIDMPNSRI
hd1-s1 TMLSEQQHSGYGVVGADQAASMTAGVSAYTDSISNSISFSSMEAGIVPDSTVIDMPNSRI
hd1-s2 TMLSEQQHSGYGVVGADQAASMTAGVSAYTDSISNSGL*
hd1-s3 TMLSEQQHSGYGVVGADQAASMTAGVSAYTDSISNSGL*
301 360
Hd1-s1 LTPAGAINLFSGPSLQMSLHFSSMDREARVLRYREKKKARKFEKTIRYETRKAYAEARPR
hd1-s1 LTPAGAINLFSGPSLQMSLHFSSMDREARVLRYREKKKARKFEKTIRYETRKAYAEARPR
hd1-s2
hd1-s3
361
Hd1-s1 IKGRFAKRSDVQIEVDQMFSTAALSDGSYGTVPWF*
hd1-s1 IKGRFAKRSDVQIEVDQMFSTAALSDGSYGTVPWF*
hd1-s2
hd1-s3

EMS_Fig. 5


1
60
Rurm1-s1 MHLTLEFGGGLELLLEKSTKVHKVDLQPNDGDGKVVMKGLLAWVKSNLIKERPEMFLKGD
rurm-s1 MHLTLEFGGGLELLLEKSTKVHKVDLQPNDGDGKVVMKGLLAWVKSNLIKERPEMFLKGD
rurm1-s2 MHLTLEFGGGLELLLEKSTKVHKVDLQPNDGDGKVVMKGLLAWVKSNLIKERPEMFLKGD
rurm1-s3 MHLTLEFGGGLELLLEKSTKVHKVDLQPNDGDGKVVMKGLLAWVKSNLIKERPEMFLKGD
rurm1-s4 MHLTLEFGGGLELLLEKSTKVHKVDLQPNDGDGKVVMKGLLAWVKSNLIKERPEMFLKGD
61
Rurm1-s1 SVRPGVLVLINDCDWELCGGLDAELEEKDVVVFISTLHGG*
rurm-s1 SVRPGVLVLIRPVTMGVSLVCHAHLIGVRLNKK*
rurm1-s2 SVRPGVLVLIRPVTMGVSLVCHAHLIGVRLNKK*
rurm1-s3 SVRPGVLVLIRPVTMGVSLVCHAHLIGVRLNKK*
rurm1-s4 SVYDFPFPSHGLCA*

EMS_Fig. 6


1 60
Ehd1-s1 MDHRELWPYGLRVLVIDDDCSYLSVMEDLLLKCSYKVTTYKNVREAVPFILDNPQIVDLV
Ehd1-s2 MDHRELWPYGLRVLVIDDDCSYLSVMEDLLLKCSYKVTTYKNVREAVPFILDNPQIVDLV
Ehd1-s3 MDHRELWPYGLRVLVIDEDCSYLSVMEDLFLKGSYKVTTYKNVREGAPFILDNPQIVDLV
ehd1-s1 MDHRELWPYGLRVLVIDDDCSYLSVMEDLLLKCSYKVAGIICTQNPARRGRMRRSHGGGG
ehd1-s2 MDHRELWPYGLRVLVIDDDCSYLSVMEDLLLKCSYKVMASSGDTNTVMKYVANGAFDFLL
ehd1-s3 MDHRELWPYGLRVLVIDDDCSYLSVMEDLLLKCSYKGKIPSIQDT*
ehd1-s4 MDHRELWPYGLRVLVIDDDCSYLSVMEDLLLKCSYKANRHMEYSQR*
ehd1-s5 MDHRELWPYGLRVLVIDDDCSYLSVMEDLLLKCSYKANRHMEYSQR*
ehd1-s6 MDHRELWPYGLRVLVIDDDCSYLSVMEDLLLKCSYKANRHMEYSQR*
ehd1-s7 MDHRELWPYGLRVLVIDDDCSYLSVMEDLLLKCSYKANRHMEYSQR*
ehd1-s8 MDHRELWPYGLRVLVIDDDCSYLSVMEDLLLKCSYKANRHMEYSQR*
61 120
Ehd1-s1 ISDAFFPTEDGLLILQEVTSKFGIPTVIMASSGDTNTVMKYVANGAFDFLLKPVRIEELS
Ehd1-s2 ISDAFFPTEDGLLILQEVTSKFGIPTVSK*
Ehd1-s3 IRDAFFPTEDGLLILQEVTSKFGIPTVSK*
ehd1-s1 RKRSGE*
ehd1-s2 KPVRIEELSNIWQHIFRKQMQDHKNNNMVGNLEKPGHPPSILAMARATPATTRSTATEAS
ehd1-s3
ehd1-s4
ehd1-s5
ehd1-s6
ehd1-s7
ehd1-s8
121 180
Ehd1-s1 NIWQHIFRKQMQDHKNNNMVGNLEKPGHPPSILAMARATPATTRSTATEASLAPLENEVR
Ehd1-s2
Ehd1-s3
ehd1-s1
ehd1-s2 LAPLENEVRDDMVNYNGEITDIRDLGKSRLTWTTQLHRQFIAAVNHLGEDKAVPKKILGI
ehd1-s3
ehd1-s4
ehd1-s5
ehd1-s6
ehd1-s7
ehd1-s8
181 240
Ehd1-s1 DDMVNYNGEITDIRDLGKSRLTWTTQLHRQFIAAVNHLGEDKAVPKKILGIMKVKHLTRE
Ehd1-s2
Ehd1-s3
ehd1-s1
ehd1-s2 MKVKHLTREQVASHLQKYRMRLKKSIPTTSKHGATLSSTALDKTQDHPSRSQYFNQDGCK
ehd1-s3
ehd1-s4
ehd1-s5
ehd1-s6
ehd1-s7
ehd1-s8
241 300
Ehd1-s1 QVASHLQKYRMQLKKSIPTTSKHGATLSSTALDKTQDHPSRSQYFNQDGCKEIMDYSLPR
Ehd1-s2
Ehd1-s3
ehd1-s1
ehd1-s2 EIMVYSLPRDDLSSDSECMLEELNDYSSEGFQDFRWDSDKQEYGPCFWNF*
ehd1-s3
ehd1-s4
ehd1-s5
ehd1-s6
ehd1-s7
ehd1-s8
301
Ehd1-s1 DDLSSGSECMLEELNDYSSEGFQDFRWDSDKQEYGPCFWNF*
Ehd1-s2
Ehd1-s3
ehd1-s1
ehd1-s2
ehd1-s3
ehd1-s4
ehd1-s5
ehd1-s6
ehd1-s7
ehd1-s8

EMS_Fig. 7