Accidental Amplification and Inactivation of a Methyltransferase Gene Eliminates Cytosine Methylation in *Mycosphaerella graminicola*

Braham Dhillon,^{*,1} Jessica R. Cavaletto,[†] Karl V. Wood[‡] and Stephen B. Goodwin^{†,2}

^{*}Department of Botany and Plant Pathology and [†]Department of Chemistry, Purdue University, West Lafayette, Indiana 47907
and [‡]U.S. Department of Agriculture–Agricultural Research Service (USDA–ARS), Crop Production and Pest Control
Research Unit, Purdue University, West Lafayette, Indiana 47907

Manuscript received April 6, 2010
Accepted for publication July 3, 2010

ABSTRACT

A *de novo* search for repetitive elements in the genome sequence of the wheat pathogen *Mycosphaerella graminicola* identified a family of repeats containing a DNA cytosine methyltransferase sequence (MgDNMT). All 23 MgDNMT sequences identified carried signatures of repeat induced point mutation (RIP). All copies were subtelomeric in location except for one on chromosome 6. Synteny with *M. fijiensis* implied that the nontelomeric copy on chromosome 6 served as a template for subsequent amplifications. Southern analysis revealed that the MgDNMT sequence also was amplified in 15 additional *M. graminicola* isolates from various geographical regions. However, this amplification event was specific to *M. graminicola*; a search for MgDNMT homologs identified only a single, unmutated copy in the genomes of 11 other ascomycetes. A genome-wide methylation assay revealed that *M. graminicola* lacks cytosine methylation, as expected if its MgDNMT gene is inactivated. Methylation was present in several other species tested, including the closest known relatives of *M. graminicola*, species S1 and S2. Therefore, the observed changes most likely occurred within the past 10,500 years since the divergence between *M. graminicola* and S1. Our data indicate that the recent amplification of a single-copy MgDNMT gene made it susceptible to RIP, resulting in complete loss of cytosine methylation in *M. graminicola*.

MYCOSPHAERELLA *graminicola* (anamorph: *Septoria tritici*) is the causal organism of *Septoria tritici* blotch (STB), which is one of the most important diseases of wheat worldwide (Eyal *et al.* 1985). In addition to Europe and the United States, infection is very severe in other wheat-growing countries in the Mediterranean, East Africa, and Australia, causing significant reductions in yield and quality. Control of STB is compounded by the development of resistance in *M. graminicola* to the benzimidazoles (Fisher and Griffin 1984) and strobilurins (Fraaije *et al.* 2005), two classes of fungicide with different modes of action.

Knowledge about pathogen biology, epidemiology, and other related physiological processes could facilitate the design of better strategies for pathogen control and reduction of disease losses. A key resource to better understand these processes is the availability of a genome sequence. Improvements in sequencing technologies and the relatively small sizes of fungal genomes

have facilitated a substantial increase in the number of species sequenced during the last few years.

An important feature found in most sequenced genomes (IHGSC 2001; Galagan *et al.* 2003) is the presence of transposable elements (TEs), DNA fragments that can move to new locations in the host genome. Autonomous TEs code for proteins for their own movement and also can mobilize similar non-autonomous elements that have lost their protein-coding capacity. Due to their ability to mobilize and increase their copy number, TEs may occupy up to 21% of some fungal genomes (Martin *et al.* 2008). TEs may be beneficial or detrimental to their host genomes. For example, in *Drosophila*, telomeres are maintained by HeT-A and TART retrotransposons (Pardue and Debaryshe 2003). Alternatively, insertions of TEs into genes can eliminate gene function or lead to the production of chimeric or antisense transcripts, thereby modifying gene expression (Feschotte 2008). In addition to insertional mutagenesis, TEs can contribute to chromosomal rearrangements by facilitating ectopic recombination (Wessler 2006). To minimize their deleterious effects, host genomes have evolved various strategies to minimize the movement of TEs, one of which involves DNA methylation.

DNA cytosine methylation occurs in many well-studied eukaryotes, with a few model organisms, such as the yeasts *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*,

Supporting information is available online at <http://www.genetics.org/cgi/content/full/genetics.110.117408/DC1>.

¹Present address: Department of Forest Sciences, University of British Columbia, Forest Sciences Centre No. 3032, 2424 Main Mall, Vancouver, British Columbia, Canada V6T 1Z4.

²Corresponding author: USDA–ARS, Crop Production and Pest Control Research Unit, Department of Botany and Plant Pathology, 915 W. State St., Purdue University, West Lafayette, IN 47907-2054.
E-mail: sgoodwin@purdue.edu

and the nematode *Caenorhabditis elegans* being the notable exceptions (COLOT and ROSSIGNOL 1999). DNA methylation is catalyzed by a conserved set of proteins called DNA methyltransferases (DNMTs), which usually add a methyl group to cytosine. DNA methylation in plants, mammals, and the filamentous fungus *Neurospora crassa* is mostly concentrated on repetitive elements (RABINOWICZ *et al.* 1999) and, at least in some cases, limits their movement. Conversely, genome-wide demethylation in plants has been shown to increase the rate of TE activity (MIURA *et al.* 2001; SINGER *et al.* 2001). Numerous qualitative and quantitative methods are available to measure DNA methylation, either for a specific locus or on a global scale. Liquid chromatography coupled with electrospray ionization tandem mass spectrometry (ESI-MS/MS) has been used to determine the whole-genome methylation status of human cell lines with very high sensitivity (SONG *et al.* 2005).

Another genome-defense mechanism, which is strictly limited to fungi, is repeat induced point mutation (RIP). RIP, discovered in *N. crassa*, was the first genome-defense system to be described in eukaryotes (SELKER *et al.* 1987; GALAGAN and SELKER 2004). During the sexual phase, RIP introduces C:G to T:A mutations in both copies of duplicated sequences >400 bp (WATTERS *et al.* 1999). These transition mutations often introduce stop codons, thus inactivating gene expression. Linked, as well as unlinked, dispersed repetitive sequences, including genes, are equally likely to be inactivated, although their genomic location may influence their susceptibility to RIP. In *N. crassa*, an ectopic insertion led to the duplication of an otherwise single-copy gene, *cya-8* (cytochrome aa3 deficient), which then was targeted and mutated by RIP (PERKINS *et al.* 2007). Up to 30% of C:G pairs in duplicated sequences can be mutated during a single sexual cycle (CAMBARERI *et al.* 1989). RIP-mutated sequences are frequently methylated in *N. crassa* (GALAGAN *et al.* 2003).

In *N. crassa*, one known DNMT, DIM-2 (defective in methylation-2) (FOSS *et al.* 1993), and one putative DNMT, RID (RIP defective) (FREITAG *et al.* 2002), have been characterized genetically. All apparent DNA methylation is lost in *dim-2* mutants without causing growth defects (FOSS *et al.* 1993), whereas mutations in the *rid* gene eliminate RIP completely. RIP has been characterized extensively in *N. crassa* but no report is available for *M. graminicola*. Although direct evidence is lacking in *M. graminicola*, repetitive sequences show characteristics of RIP, such as high ratios of transitions to transversions and low GC content (GALAGAN and SELKER 2004).

Due to its economic importance and genetic tractability, the genome of *M. graminicola* was sequenced to completion by the Joint Genome Institute (JGI) of the U.S. Department of Energy (<http://genome.jgi-psf.org/Mycgr3/Mycgr3.home.html>). It is the first genome of a filamentous fungus to be finished according to recently proposed standards (CHAIN *et al.* 2009); *i.e.*, all 21

chromosomes except the smallest one have been sequenced completely from telomere to telomere with only two gaps of unclonable DNA. We characterized the repetitive sequences in the finished *M. graminicola* genome to search specifically for high-copy-number gene families as potential targets of RIP. Our aim was to test whether genes or gene fragments have been amplified in the genome and how this amplification could be influenced by RIP. We found an example where a single-copy DNMT gene had been amplified in the *M. graminicola* genome. This gene-amplification event was most likely followed by RIP-mediated gene inactivation and a concomitant species-wide loss of methylation in *M. graminicola*.

MATERIALS AND METHODS

Identification and characterization of repetitive sequences:

Sequences of two closely related fungi, the wheat pathogen *M. graminicola* and the banana pathogen *M. fijiensis*, were used in this study. The 21 chromosomes of *M. graminicola* isolate IPO323 have been sequenced completely (JGI, v.2.0, <http://genome.jgi-psf.org/Mycgr3/Mycgr3.home.html>) while for *M. fijiensis*, a 7.1× draft sequence is available (JGI, v.1.0, <http://genome.jgi-psf.org/Mycfi1/Mycfi1.home.html>).

RECON version 1.05 (BAO and EDDY 2002) was used to identify repetitive elements *de novo* in the *M. graminicola* genome. During characterization of these repetitive sequences, a repeat family was identified, designated as “fam1,” members of which contained a protein sequence similar to a DNA methyltransferase (MgDNMT). Fam1 elements were further searched for protein domains generally associated with repetitive elements, using BLASTX (ALTSCHUL *et al.* 1997) against the NCBI nonredundant protein database (nr) at e^{-5} , and for structural features such as terminal inverted repeats (TIR) using “Einverted” (<http://emboss.sourceforge.net/apps/release/6.2/emboss/apps/einverted.html>) and tandem repeats using tandem repeat finder (TRF) version 4.0 (BENSON 1999).

Identification and phylogenetic analysis of Dim2-like genes in other fungal genomes: Sequences similar to the *N. crassa* DIM-2 gene were identified and copy number was determined in 10 other fungal species: *Botryotinia fuckeliana*, *Coccidioides immitis*, *Gibberella zeae*, *Magnaporthe oryzae*, *Phaeosphaeria nodorum*, *Pyrenophora tritici-repentis*, *Sclerotinia sclerotiorum*, *Podospora anserina*, *M. fijiensis*, and *M. graminicola* relative S1. The genome sequences for these species were available at <http://www.broad.mit.edu/annotation/fgi/>, <http://podospora.igmors.u-psud.fr/> and <http://genome.jgi-psf.org/Mycfi1/Mycfi1.home.html>. DIM-2-like sequence from the *M. graminicola* relative S1 was provided by E. Stukenbrock. Including the Dim-2 gene from *N. crassa* and the “RIPed” and “deRIPed” *M. graminicola* sequences, 13 DIM-2-like sequences were aligned using ClustalX version 2.0 (THOMPSON *et al.* 1997) and a neighbor-joining (SAITOU and NEI 1987) tree was constructed using a 100-replicate bootstrap analysis.

RIP and deRIP in MgDNMT: To quantify the transitions induced by RIP, DNA sequences of all fam1 elements were aligned using ClustalX version 2.0 (THOMPSON *et al.* 1997) and were edited manually using Jalview version 2.3 (CLAMP *et al.* 2004). DeRIPing was done by comparing the base composition at each nucleotide position of aligned sequences and deducing the original sequence prior to RIP. For this analysis, the only nontelomeric MgDNMT sequence, on chromosome 6,

was used as a reference. At each polymorphic base position in the alignment, any thymine (T) residue in the reference sequence was changed to a cytosine (C), if any of the other sequences had a corresponding C. Similarly, an adenine (A) in the reference sequence was changed to a guanine (G) when a "G" was present among other sequences in the alignment. This corrected for RIP on both strands of the DNA sequence. Remaining internal stop codons were corrected manually by inserting transitions that changed the stop to an amino acid, usually glutamine (Q), when compared to the S1 DNMT copy. These stops were assumed to be at sites that were RIPed in all the copies and were no longer polymorphic.

Southern hybridization for estimation of copy number: Sixteen isolates were used for this analysis. Fifteen isolates of *M. graminicola* from seven countries on five continents, including Turkey (isolates IPO86013 and IPO86022), Argentina (IPO86068), Uruguay (IPO87016, IPO87019), Ethiopia (IPO88004, IPO88018), The Netherlands (IPO001, IPO235, IPO89011, IPO323), Australia (Paskeville), and the United States (IIA.1, IIA.3, ST2), and one isolate of *S. passerinii* (P77) were used. These isolates were grown in liquid culture as described previously (GOODWIN *et al.* 2001). Mycelia were collected and lyophilized before DNA extraction (DNeasy Plant Mini Kit; QIAGEN, Valencia, CA). Southern analysis was done using alkaline transfer (SAMBROOK and RUSSELL 1989), and chemiluminescence was used to detect the DIG-labeled MgDNMT probe (Roche Diagnostics, Indianapolis). The probe was designed from the nontelomeric MgDNMT reference sequence on chromosome 6.

Determination of syntenic regions for MgDNMT: To determine synteny, 50 kb of sequence flanking the MgDNMT sequence was compared to the homologous sequences in *M. fijiensis*, *P. tritici-repentis*, *P. nodorum*, and *S. sclerotiorum*. Graphic visualization of genomic comparisons was done using the Artemis Comparison Tool (ACT) Release 8 (CARVER *et al.* 2005). This tool was used to display similarity at the nucleotide level even though at the amino acid level the similarity was much higher.

Analysis of genome-wide cytosine methylation: Three isolates of *M. graminicola* from different geographical regions (IPO323, IPO86068, Paskeville), four isolates from two new, undescribed species of *Mycosphaerella* from wild grasses (species S1, isolates ST04IR-221 and ST04IR-431; and S2, isolates ST04IR-111 and ST04IR-3131), three isolates of *M. fijiensis* (IPO139a, IPO8837, rCRB2), and one isolate of *S. passerinii* (P63) were used. DNA was hydrolyzed as described previously (SONG *et al.* 2005). This mixture was analyzed by electrospray ionization tandem mass spectrometry. Briefly, 5 μ g of genomic DNA was denatured by heating to 100° for 3 min before chilling on ice, followed by addition of 1/10 vol of 0.1 M ammonium acetate (pH 5.3) and 2 units of nuclease P1 and incubation at 45° for 2 hr. Next, 1/10 vol of 1 M ammonium bicarbonate and 0.002 unit of venom phosphodiesterase I were added and the mixture was incubated at 37° for 2 hr. The final incubation was at 37° for 1 hr after addition of 0.5 unit of alkaline phosphatase.

All ESI analyses were carried out on a Finnigan MAT LCQ Classic mass spectrometer system (ThermoElectron, San Jose, CA). The electrospray needle voltage and the heated capillary voltage were set to 4.0 kV and 10 V, respectively. The capillary temperature was set at 207° and the typical background source pressure was 1.2×10^{-5} torr. The sample flow rate was ~ 8 μ l per minute. The drying gas was nitrogen. The LCQ was scanned to 1000 amu for these experiments. The sample was dissolved in methanol and water.

The MS/MS results were obtained by selecting the ion of interest (the precursor ion). The precursor ion was then subjected to collision-induced dissociation (CID) resulting in

the formation of product ions. Helium was introduced into the system to an estimated pressure of 1 mtorr to improve trapping efficiency and also acted as the collision gas during the CID experiments. The collision energy was set to 40% of the maximum available from the 5-V tickle voltage, with a 2-mass-unit isolation window.

RESULTS

Characterization of fam1 repeats: Repeat analysis of the finished sequence of the *M. graminicola* genome (<http://genome.jgi-psf.org/Mycgr3/Mycgr3.home.html>) with RECON (BAO and EDDY 2002) identified 106 families of repetitive elements. One of these, family 1 (fam1) had a total of 16 elements on seven chromosomes (supporting information, Figure S1) with the longest element, ele9870 at ~ 9.7 kb long. All fam1 elements except for one on chromosome 6 were located in subtelomeric regions. No other distinguishing structural features associated with TEs such as direct or inverted repeats were found in the elements of this family.

A search for protein domains in fam1 elements identified a DNA cytosine methyltransferase protein (MgDNMT), which is atypical for repetitive elements. This MgDNMT belongs to a class of proteins represented by DIM-2, a genetically well-characterized protein from *N. crassa*. After searching the *M. graminicola* genome with the *N. crassa* DIM-2 (NcDim-2) sequence, 12 additional copies of the MgDNMT-containing repeats were identified, also at subtelomeric locations. Of the 28 fam1 elements distributed on 17 chromosomes, the MgDNMT region was present in 23 (Table S1). However, all elements lacked the usual protein-coding regions commonly associated with TEs such as transposases, reverse transcriptases, or helicases.

Determining the original MgDNMT region and synteny with other genomes: To determine how the number of MgDNMT sequences increased in the genome, it was essential to identify the original MgDNMT sequence. One interesting candidate for the original "donor" sequence was the 4.4-kb nontelomeric copy present on chromosome 6. Sequences flanking the element on chromosome 6 were unique, suggesting that this 4.4-kb sequence is the possible source for the MgDNMT region in fam1 elements. This comparison helped delimit the extent of the genomic fragment that had become a part of the fam1 repetitive elements.

A search of the genome of *M. fijiensis*, a relative of *M. graminicola*, using the MgDNMT sequence identified a single-copy homolog on scaffold 1. Fifty kilobases of genomic sequence flanking all fam1 elements was used to search the *M. fijiensis* scaffold 1 sequence to identify possible syntenic regions. The only region in *M. graminicola* conserved for gene number and order between the two genomes was the nontelomeric region on chromosome 6. Two genes present upstream of the MgDNMT sequence, one similar to a UsgS transmem-

FIGURE 1.—Synteny between *Mycosphaerella graminicola* and *M. fijiensis* in the MgDNMT region. One hundred kilobases of sequence on chromosome 6 from *M. graminicola* and scaffold 1 from *M. fijiensis* were compared to test for synteny of gene content and order between the two genomes. Graphs at the top and bottom show the percentage of GC for *M. graminicola* and *M. fijiensis*, respectively. Positions and orientations of genes are indicated in the tracks for each region. Syntenic genes are linked by lines indicating similarity at the nucleotide level. The shaded regions in the percentage of GC panels show a sharp decrease in GC content in the MgDNMT region of *M. graminicola* as compared to *M. fijiensis*.

brane protein and the other similar to a conserved hypothetical protein, were present in *M. fijiensis* but in reverse orientation (Figure 1). In *M. graminicola* these genes are ~0.4 kb apart, whereas in *M. fijiensis* they are separated by a 4.4-kb repetitive element. Similarly, downstream of the MgDNMT sequence is a C6 zinc finger domain-containing protein that is present at the same location in *M. fijiensis* (Figure 1). Therefore, the four genes in this region appear to be syntenic between *M. graminicola* and *M. fijiensis* but not collinear. All three genes surrounding the MgDNMT sequence are specific to fungi.

The four genes shared between *M. graminicola* and *M. fijiensis* were present in *P. tritici-repentis*, *P. nodorum*, and *S. sclerotiorum* but were not syntenic. This synteny between *M. fijiensis* scaffold 1 and the MgDNMT sequence on *M. graminicola* chromosome 6 strengthens the possibility that the nontelomeric copy on chromosome 6 was the original template for subsequent amplifications.

Amplification of MgDNMT in *M. graminicola*: Only one match to the NcDIM-2 protein was found in the genome sequences of 11 phylogenetically diverse species (Figure 2) compared to the 23 copies identified in *M. graminicola*. The RIPed copy of the MgDNMT sequence was on a much longer branch reflecting its greater rate of change compared to the deRIPed version. Southern analysis using the chromosome 6 MgDNMT sequence as probe revealed multiple copies in 15 *M. graminicola* isolates from various geographical regions

(Figure 3). Higher intensity of some bands, especially in lanes 1 (IPO86013) and 2 (IPO86022), may represent multiple copies, which were either similar in size or too large to be resolved on the gel (Figure 3). Lanes 3 (IPO86068) and 10 (IIA.1) contained much less DNA and showed either no or very little hybridization. However, an overexposure revealed that these isolates had at least five bands. No hybridization with the MgDNMT sequence was visible in a closely related fungus, the barley pathogen *S. passerinii*, even after overexposure.

Subtelomeric location of fam1 elements: Each fam1 element contained sequences similar to one or both of the two longest elements, ele799 or ele9870 (Figure S1). On two pairs of chromosomes, 2 and 17 (Figure 4) and 9 and 21, fam1 elements were duplicated and formed complex repetitive structures at the subtelomeric locations. Comparisons of these regions revealed a distinct arrangement of ele799-like and ele9870-like elements with the MgDNMT region sandwiched between them.

Pairwise comparisons of 23 MgDNMT sequences pointed at one possible mechanism of MgDNMT sequence amplification in the genome. Two nearly identical sequences (2.4 kb each with 99% identity) were found at the same chromosomal position, adjacent to the telomeric repeats, one each on chromosomes 9 and 21. The two sequences differ by seven transitions and one transversion mutation, of which six caused synonymous amino acid substitutions. Other MgDNMT se-

FIGURE 2.—Phylogenetic relationships among DIM-2-like genes in *Mycosphaerella graminicola* and 11 other ascomycete species. A neighbor-joining tree is shown for 9 phylogenetically diverse species, along with RIPed and deRIPed *M. graminicola* sequence, *M. fijiensis*, and species S1, using 100 bootstrap replicates. Sequences corresponding to the species names are as follows: CIMG_01762, *Coccidioides immitis*; PODANSg1750, *Podospira anserina*; FG10766, *Gibberella zeae*; BC1G_12419, *Botryotinia fuckeliana*; MGG_00889, *Magnaporthe oryzae*; SNOG_03039, *Phaeosphaeria nodorum*; PTRG_02280, *Pyrenophora tritici-repentis*; SS1G_07976, *Sclerotinia sclerotiorum*; and NCU02247, *Neurospora crassa*. Bootstrap values were $\geq 85\%$ and are indicated at the nodes.

quences showed transition mutations when compared to these two. A 99% identity between 2 MgDNMT copies and significant differences from the other 21 copies implies that amplification may have occurred recently, subsequent to mutagenesis by RIP. The most likely explanation is that after accumulating mutations, one sequence could have been copied to a new location via reciprocal translocation between the subtelomeres.

Juxtaposition of an LTR retrotransposon fragment:

Repetitive sequences flanking the MgDNMT region in fam1 elements were devoid of any repeat-associated protein domains or characteristic structural features commonly associated with TEs. To test for other similarities to TEs, these flanking sequences were used to search the repetitive sequence library from *M. graminicola*. A 220-bp fragment immediately upstream of the MgDNMT sequence was similar to a noncoding region from a long terminal repeat (LTR) retrotransposon family (fam15) (Figure S2). In the LTR retrotransposon sequence, the 220-bp fragment was present immediately upstream of the 3' LTR. This suggests that at some subtelomeric location, MgDNMT sequence was adjacent to a LTR retrotransposon. This similarity also hints at the possibility that the MgDNMT sequence might have been acquired by a LTR retrotransposon and then moved to a telomeric location.

RIP and deRIP in MgDNMT: Multiple copies of the MgDNMT sequence in the *M. graminicola* genome make it a likely candidate for RIP. To determine the extent of RIP, the original chromosome 6 MgDNMT sequence was

deRIPed by replacing a T with a C, or an A with a G, at positions showing C/T or A/G polymorphisms, respectively, and compared to the remaining MgDNMT copies. The number of transversion mutations between the MgDNMT copies and the deRIPed version of MgDNMT sequence ranged from 0 to 11 per kilobase of sequence compared (Table 1). Two subsets of sequences were present, one with greater than four transversions per kilobase (15 sequences) and the other with fewer than two transversions per kilobase (7 sequences). The sequences with the higher numbers of transversions included all of the full-length MgDNMT sequences, whereas in the other set all sequences were truncated.

Transition mutations varied from 14 to 107 per kilobase of sequence analyzed. These transitions resulted in 61 stop codons in the putative coding region of the nontelomeric MgDNMT sequence (Figure S3). All fam1 elements, except two, had stop codons that were caused by C $\rightarrow$ T and G $\rightarrow$ A transitions. Because the two remaining fam1 elements were truncated, no functional copy of the MgDNMT sequence appears to be present in the *M. graminicola* genome. Moreover, no sequences corresponding to MgDNMT were identified in the *M. graminicola* EST data set. In contrast, the single-copy *dim-2* homologs in the genomes of *M. fijiensis* and the other fungal species analyzed were complete with no evidence of RIP or stop codons.

Searching the genome sequence of species S1, the closest known relative of *M. graminicola* (STUKENBROCK *et al.* 2007) with the deRIPed MgDNMT copy along with

FIGURE 3.—Southern analysis of MgDNMT in *Mycosphaerella graminicola*. The MgDNMT probe highlights a number of bands in each lane. Lanes 3 and 10 show weak hybridization, but this was mostly due to unequal loading and an overexposure revealed multiple bands. Geographic regions and isolates in order of loading are as follows: Turkey, IPO86013 and IPO86022; Argentina, IPO86068; Uruguay, IPO87016 and IPO87019; Ethiopia, IPO88004 and IPO88018; The Netherlands, IPO001 and IPO235; United States, IIA.1 and ST2; Australia, Paskeville; and United States, IIA.3. Also tested but not shown were IPO89011 (The Netherlands), IPO323 (Netherlands isolate that was sequenced), and *S. passerinii* P77. A 1-kb ladder (lane M) was used as a size standard. The scale at the bottom shows the MgDNMT region from which the probe was designed. The arrows mark the flanking *Pst*I restriction enzyme sites.

its flanking sequence identified only one copy of the MgDNMT-like sequence. On the basis of comparisons to its putative homolog in S1, the 4.4-kb chromosome 6 donor sequence most likely contains the complete predicted DNMT sequence coding for 1281 amino acids.

After adjusting the nucleotide polymorphisms for RIP, the chromosome 6 MgDNMT sequence still had eight stop codons that were resolved by comparison to the S1 sequence. This final deRIPed MgDNMT sequence showed an improvement in significance value of BLAST searches to the NCBI “nr” database from e^{-52} to e^{-175} . The expected translation product of the deRIPed MgDNMT sequence shows significant similarity to other fungal DIM-2-like proteins, but the N-terminal and C-terminal ends were of different lengths and did not align (Figure S4).

Recent loss of cytosine methylation in *M. graminicola*: *N. crassa* DIM-2, a MgDNMT homolog, is responsible for all of the known cytosine methylation in that organism (KOUZMINOVA and SELKER 2001). To determine whether RIP of the MgDNMT sequence affected DNA methylation of *M. graminicola*, a genome-wide DNA methylation assay was conducted using ESI-MS/MS.

Under the assay conditions, the presence of a specific ion can be confirmed from its fragmentation products;

i.e., loss of the dehydrated deoxyribose sugar residue from 5-methyl deoxycytidine (5mdC) will give rise to 5-methyl cytosine (5mC). This can be detected qualitatively in ESI-MS/MS spectra by the presence of an ion at mass-to-charge ratio (m/z) 126 (after loss of dehydrated deoxyribose), in the MS/MS spectrum of m/z 242 (protonated 5mdC). The control MS/MS spectra of m/z 228, protonated cytidine (dC), for both *M. fijiensis* (Figure 5A) and *M. graminicola* (Figure 5B) show a similar fragment ion at m/z 112 indicating a loss of dehydrated deoxyribose. However, differences were observed in the MS/MS spectra of m/z 242. In *M. fijiensis* (Figure 5C), the fragment ion at m/z 126 (5mC) is clearly visible. However, this fragmentation ion is absent in the MS/MS spectrum of m/z 242 from *M. graminicola* (Figure 5D). Therefore, as compared to *M. fijiensis*, cytosine methylation is absent from *M. graminicola*. The cytosine methylation profile of the closely related barley pathogen *S. passerinii* was similar to that for *M. fijiensis*.

Isolates of the recently discovered, unnamed species S1 and S2 from uncultivated grasses in Iran also were assayed for cytosine methylation. Cytosine methylation was present in both S1 and S2, which are thought to have diverged from *M. graminicola* ~10,500 and 20,000 years ago, respectively (STUKENBROCK *et al.* 2007). Therefore, loss of cytosine methylation in *M. graminicola* probably happened after its divergence from S1 within the past 10,500 years.

DISCUSSION

Multiple copies of a DNA methyltransferase (MgDNMT) sequence at subtelomeric positions in the *M. graminicola* genome, all marred with signatures of RIP, is unprecedented. A DNMT domain is not a component of known repetitive elements but, due to its high copy number, the MgDNMT sequence was recognized *de novo* as a repetitive sequence by RECON. The species-wide presence of this amplification event in *M. graminicola* was suggested by the occurrence of multiple copies of MgDNMT sequence in 15 isolates from diverse geographical regions. However, only one copy was present in the genome sequence of the closely related species S1 and 10 other ascomycete fungi, so the expansion seems to be recent and unique to *M. graminicola*.

RIP protects fungal genomes from the detrimental effects of repetitive elements (GALAGAN and SELKER 2004). The only gene known to be involved in RIP is the RID gene from *N. crassa*, which also is a putative DNMT (FREITAG *et al.* 2002). However, the exact role that RID plays in RIP is unknown. DIM-2, on the other hand, is not required for RIP (KOUZMINOVA and SELKER 2001; FREITAG *et al.* 2002). The genomic sequence of *M. graminicola* contains a putative ortholog of the *rid* gene. It is single copy and predicted to encode a protein of 742

FIGURE 4.—Comparison of fam1 subtelomeric repeats on chromosomes 2 and 17 of *Mycosphaerella graminicola*. Similar sequences on chromosomes 2 and 17 are connected by straight (A) or slanting (B) lines. Locations of ele799- and ele9870-like elements and the MgDNMT sequence are indicated by boxes with dark shading, hatch marks, and horizontal lines, respectively. Telomeres are indicated by circles. Chromosome positions are indicated in kilobases. GC content plots are shown above and below each chromosome.

amino acids, suggesting that RIP is functional in *M. graminicola*.

Another feature of RIP-induced transitions is a decrease in GC content of the affected sequences. A

corresponding decrease in percentage of GC was notable in the DNMT region in *M. graminicola* as compared to *M. fijiensis* (Figure 1). Substituting the nucleotides A/T with G/C at polymorphic sites of aligned

TABLE 1
Pairwise comparison of MgDNMT sequences to the deRIPed nontelomeric sequence on chromosome 6 of *Mycosphaerella graminicola*

No.	Chromosome	Element	Length	Ti ^a	Tv ^a	Ti/Tv	Ti/kb	Tv/kb
1	8	1_24	1710	174	19	9.2	101.8	11.1
2	2	1_08	4419	330	29	11.4	74.7	6.6
3	2	1_09	4419	331	29	11.4	74.9	6.6
4	17	1_11	4422	306	26	11.8	69.2	5.9
5	17	1_12	4416	176	25	7.0	39.9	5.7
6	5	1_14	4419	247	25	9.9	55.9	5.7
7	1	1_02	4422	467	25	18.7	105.6	5.7
8	15	1_10	4422	309	24	12.9	69.9	5.4
9	21	1_16	4422	279	24	11.6	63.1	5.4
10	1	1_01	4422	472	23	20.5	106.7	5.2
11	1	1_03	4076	345	21	16.4	84.6	5.2
12	1	1_04	4078	341	21	16.2	83.6	5.1
13	6	1_18	4419	227	22	10.3	51.4	5.0
14	5	1_15	4244	327	21	15.6	77.0	4.9
15	1	1_06	453	10	2	5.0	22.1	4.4
16	12	1_21	2271	202	3	67.3	88.9	1.3
17	9	1_23	2429	37	3	12.3	15.2	1.2
18	21	1_17	2429	34	2	17.0	14.0	0.8
19	18	1_25	1478	109	1	109.0	73.7	0.7
20	1	1_05	2221	75	1	75.0	33.8	0.5
21	17	1_13	2564	53	1	53.0	20.7	0.4
23	16	1_22	2136	109	0	—	51.0	0.0
24 ^b	6	1_19_20	4331	435	0	—	100.4	0.0

All sequences were sorted by transversions per kilobase in decreasing order.

^a Ti, transitions; Tv, transversions.

^b The original RIPed copy on chromosome 6.

FIGURE 5.—Comparison of ESI-MS/MS mass spectra of *Mycosphaerella fijiensis* (A and C) and *M. graminicola* (B and D) genomic DNA. Product ion mass spectra, m/z 228 and m/z 242 of deoxycytidine (A and B) and 5-methyl deoxycytidine (C and D), and their fragmentation products following loss of dehydrated deoxyribose at m/z 112 and 126, respectively, are shown. At least 10 scans were averaged. Absence of the m/z 126 peak in *M. graminicola* (D) is indicated by a downward-pointing arrow. The fragment ion at m/z 242 can represent other ions, but the transition from m/z 242 to m/z 126 comes only from 5-methyl cytosine. The fragment ions at m/z 187 and 215 correspond to unknowns that are present in DNA of both species.

sequences increased the percentage of GC content of the deRIPed MgDNMT sequence from 43 to 53. Transitions introduced by RIP inactivated all MgDNMT copies in the genome, leading to a loss of cytosine methylation. The absence of cytosine methylation in *M. graminicola* was alluded to in a previous analysis based on methylation-sensitive and -insensitive restriction enzymes (GOODWIN *et al.* 2001). However, lack of cytosine methylation is not unique among fungi; many yeasts, such as *S. cerevisiae* and *S. pombe*, lack detectable methylation (COLOT and ROSSIGNOL 1999). Moreover, methylation is not essential in *N. crassa*, as *dim-2* mutants are viable. These mutants show no obvious phenotype under laboratory conditions despite genome-wide demethylation (KOUZMINOVA and SELKER 2001). Therefore, DNA methylation is not required in all organisms and the lack of a functional copy of DNMT has no obvious fitness cost in *M. graminicola*.

In the absence of DNA methylation, histone methylation might be responsible for the silencing of repetitive DNA. In *N. crassa*, the *dim-5* gene is responsible for methylation of histone H3 at the lysine 9 residue (H3K9). Cytosine methylation depends upon prior methylation of H3K9 residues (TAMARU and SELKER 2001) and binding of H3K9me by heterochromatin protein 1 (*hpo* gene) (FREITAG *et al.* 2004). Homologs

for both *dim-5* and the *hpo* gene are present in *M. graminicola*.

Identification of the original MgDNMT sequence was facilitated by synteny with *M. fijiensis*. However, synteny was nonexistent with other less closely related fungi such as *P. nodorum*, *P. tritici-repentis*, and *S. sclerotiorum*. This is not surprising, because closely related organisms typically share a higher number of similar genomic regions and synteny breaks down more drastically in comparison with distant relatives (LITI and LOUIS 2005). Comparison of orthologous regions in closely related fungi such as *N. crassa*, *F. graminearum*, and *M. grisea* reveals that syntenic blocks usually consist of small numbers of genes, ranging from 3 to 20 (XU *et al.* 2006). Even when gene order has been conserved, the transcriptional orientations of the genes relative to one another often are different, as also was observed between *M. graminicola* and *M. fijiensis*.

Repetitive sequences also can lead to loss of conserved synteny, because they promote crossing over at nonhomologous chromosomal sites leading to chromosomal rearrangements. In *S. cerevisiae*, LTR retroelements were frequently associated with rearrangements (DUNHAM *et al.* 2002). Although comprehensive synteny analysis between *M. graminicola* and *M. fijiensis* is lacking, the role of TEs in the breakdown of synteny was evident

during the comparison of 100 kb of sequence flanking the MgDNMT region between the two species. This region in *M. graminicola* had 15 predicted genes and no repetitive sequences, whereas in *M. fijiensis* 66% of the sequence was repetitive with only 4 predicted genes (Figure 1). Moreover, in *M. fijiensis*, a TE was inserted into the intergenic region of syntenic genes. The extensive presence of TEs in this region in *M. fijiensis* suggests that their insertion in intergenic regions may lead to loss of synteny between closely related species of Dothideomycetes as well.

Reverse transcriptase/endonuclease proteins, especially those in non-LTR LINE retrotransposons, can bind and transcribe cellular mRNAs and integrate them back into the genome, giving rise to processed retrogenes (ESNAULT *et al.* 2000). However, to date there appears to be only one case reported in which a LTR retrotransposon has been shown to carry gene fragments. In maize, one of the first retrotransposons to be identified was Bs1, which later was shown to carry transduced fragments from three genes (ELROUBY and BUREAU 2001). Although no mechanism for gene capture was proposed, the Bs1 results support the idea that LTR retrotransposons are capable of mobilizing and multiplying single-copy genes within genomes. The 220-bp LTR fragment adjacent to the MgDNMT region could have come from a complete LTR retroelement, which was gradually shortened during subsequent recombination events. Alternatively, this LTR fragment might have been present already at subtelomeric locations and was propagated along with the MgDNMT region.

A more plausible explanation for increase in MgDNMT copy number could be amplification by segmental duplication. One mechanism for segmental duplication is double-strand break (DSB) repair (BAILEY *et al.* 2003). Any sequence in the genome may be used ectopically as a template to initiate DSB repair, leading to the duplication of that region (RONG and GOLIC 2003). In *M. graminicola*, the original nontelomeric chromosome 6 MgDNMT sequence could have served as a template for DSB repair at a subtelomeric location. Once copied to the subtelomeric location, the MgDNMT region could have been amplified by ectopic recombination between subtelomeric regions on different chromosomes. Ectopic exchange between subtelomeres in *M. graminicola* is supported by sequences that are identical between the subtelomeric regions of at least two pairs of chromosomes. These identical regions extend well beyond the MgDNMT sequence. In *S. cerevisiae*, at least three gene families, β -fructofuranosidase, α -galactosidase, and resistance to toxicity of molasses, have been amplified between chromosome ends through ectopic recombination (LOUIS *et al.* 1994). In humans, subtelomeric regions are patches of interchromosomal segmental duplications (LINARDOPOULOU *et al.* 2005) with high plasticity,

which may increase gene diversity (TRASK *et al.* 1998), as observed in the subtelomeres of *M. graminicola*.

Four additional chromosomal ends in *M. graminicola* (between chromosomes 4 and 18 and 8 and 13) also have similar long repetitive sequences near their telomeres, but do not include MgDNMT sequence. Subtelomeric repetitive sequences have been reported in several organisms, but the reason for these structures is still unknown (FLINT *et al.* 1997). It has been suggested that in the absence of telomerase, large blocks of tandem arrays of subtelomeric repeats may help stabilize the telomeres (MCEACHERN *et al.* 2000). This mechanism of telomerase-independent, recombination-based telomere maintenance has been demonstrated in *S. cerevisiae* and *Kluyveromyces lactis* (LUNDBLAD and BLACKBURN 1993; MCEACHERN and BLACKBURN 1996) and also may operate in *M. graminicola*.

The different isolates of *M. graminicola* showed a high degree of polymorphism for bands corresponding to the MgDNMT sequence. Size polymorphisms on Southern blots usually are attributed to mutations in restriction enzyme recognition sites or the activity of transposable elements. In *M. graminicola*, two factors, RIP and high recombination at subtelomeres, may have contributed to these size polymorphisms. However, in all isolates, except for the sequenced *M. graminicola* isolate IPO323, the chromosomal location of MgDNMT sequences is unknown. Only a subset of total MgDNMT sequences was initially recognized by RECON (BAO and EDDY 2002) and the rest were identified by an iterative search. Improper definition of element boundaries is one reason that RECON (BAO and EDDY 2002) may sometimes fail to cluster similar elements into one family.

In organisms where a RIP-like process is absent, genes with altered functions may be created from transposable element-mediated multiplication of captured genes (MCCARREY and THOMAS 1987) or gene duplication and/or segmental duplication events. However, in filamentous fungi with an active RIP-like process, duplicated sequences are likely to be mutated by RIP. In *N. crassa*, only 6 from a predicted 10,082 genes have highly similar duplicates, suggesting that evolution via gene duplication has been virtually arrested (GALAGAN *et al.* 2003). Loss of function of the *N. crassa* *cya-8* gene was shown to be a direct consequence of gene duplication followed by RIP (PERKINS *et al.* 2007). Our results support the idea that accidental amplification followed by RIP may be a prevalent mechanism to inactivate single-copy genes, leading to significant effects on the basic biological pathways in fungi. These changes can occur rapidly; amplification and subsequent inactivation of MgDNMT sequences probably occurred after the split of *M. graminicola* and S1, which was estimated at 10,500 years ago (STUKENBROCK *et al.* 2007), and may have been concomitant with the domestication of wheat as a cultivated crop. Whether loss of methylation influenced

the shift in *M. graminicola* host preference from a wild grass to wheat is not known, but it provides a testable hypothesis for future research.

We thank Bruce McDonald for providing DNA samples of S1 and S2, Eva Stukenbrock for BLAST searches against the S1 genome sequence, and Gert Kema for help in obtaining the genomic sequence of *M. graminicola*. We also thank Michael Freitag and Eric Selker for critically reviewing the manuscript. DNA sequencing of *M. graminicola* and *M. fijiensis* was performed at the U.S. Department of Energy's Joint Genome Institute through the Community Sequencing Program (www.jgi.doe.gov/CSP/) and all sequence data are publicly available. This work was supported by the U.S. Department of Agriculture (USDA) Current Research Information System (CRIS) project 3602-22000-015-00D. Names are necessary to report factually on available data. However, the USDA neither guarantees nor warrants the standard of the product, and the use of the name implies no approval of the product to the exclusion of others that also may be suitable.

LITERATURE CITED

- ALTSCHUL, S. F., T. L. MADDEN, A. A. SCHAFER, J. ZHANG, Z. ZHANG *et al.*, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389–3402.
- BAILEY, J. A., G. LIU and E. E. EICHLER, 2003 An Alu transposition model for the origin and expansion of human segmental duplications. *Am. J. Hum. Genet.* **73**: 823–834.
- BAO, Z., and S. R. EDDY, 2002 Automated *de novo* identification of repeat sequence families in sequenced genomes. *Genome Res.* **12**: 1269–1276.
- BENSON, G., 1999 Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**: 573–580.
- CAMBARERI, E. B., B. C. JENSEN, E. SCHABTACH and E. U. SELKER, 1989 Repeat-induced G-C to A-T mutations in *Neurospora*. *Science* **244**: 1571–1575.
- CARVER, T. J., K. M. RUTHERFORD, M. BERRIMAN, M. A. RAJANDREAM, B. G. BARRELL *et al.*, 2005 ACT: the Artemis Comparison Tool. *Bioinformatics* **21**: 3422–3423.
- CHAIN, P. S., D. V. GRAHAM, R. S. FULTON, M. G. FITZGERALD, J. HOSTETLER *et al.*, 2009 Genomics. Genome project standards in a new era of sequencing. *Science* **326**: 236–237.
- CLAMP, M., J. CUFF, S. M. SEARLE and G. J. BARTON, 2004 The Jalview Java alignment editor. *Bioinformatics* **20**: 426–427.
- COLOT, V., and J. L. ROSSIGNOL, 1999 Eukaryotic DNA methylation as an evolutionary device. *BioEssays* **21**: 402–411.
- DUNHAM, M. J., H. BADRANE, T. FEREA, J. ADAMS, P. O. BROWN *et al.*, 2002 Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **99**: 16144–16149.
- ELROUBY, N., and T. E. BUREAU, 2001 A novel hybrid open reading frame formed by multiple cellular gene transductions by a plant long terminal repeat retroelement. *J. Biol. Chem.* **276**: 41963–41968.
- ESNAULT, C., J. MAESTRE and T. HEIDMANN, 2000 Human LINE retrotransposons generate processed pseudogenes. *Nat. Genet.* **24**: 363–367.
- EYAL, Z., A. L. SCHAREN, M. D. HUFFMAN and J. M. PRESCOTT, 1985 Global insights into virulence frequencies of *Mycosphaerella graminicola*. *Phytopathology* **75**: 1456–1462.
- FESCHOTTE, C., 2008 Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* **9**: 397–405.
- FISHER, N., and M. GRIFFIN, 1984 Benzimidazole (MBC) resistance in *Septoria tritici*. *ISPP Chem. Control Newsl.* **5**: 8–9.
- FLINT, J., G. P. BATES, K. CLARK, A. DORMAN, D. WILLINGHAM *et al.*, 1997 Sequence comparison of human and yeast telomeres identifies structurally distinct subtelomeric domains. *Hum. Mol. Genet.* **6**: 1305–1313.
- FOSS, H. M., C. J. ROBERTS, K. M. CLAEYS and E. U. SELKER, 1993 Abnormal chromosome behavior in *Neurospora* mutants defective in DNA methylation. *Science* **262**: 1737–1741.
- FRAAIJE, B. A., F. J. BURNETT, W. S. CLARK, J. MOTTERAM and J. A. LUCAS, 2005 Resistance development to QoI inhibitors in populations of *Mycosphaerella graminicola* in the UK, pp. 63–71 in *Modern Fungicides & Antifungal Compounds IV*, edited by H. W. DEHNE, U. GISI, K. H. KUCK, P. E. RUSSELL and H. LYR. BCPC, Alton, UK.
- FREITAG, M., R. L. WILLIAMS, G. O. KOTHE and E. U. SELKER, 2002 A cytosine methyltransferase homologue is essential for repeat-induced point mutation in *Neurospora crassa*. *Proc. Natl. Acad. Sci. USA* **99**: 8802–8807.
- FREITAG, M., P. C. HICKEY, T. K. KHLAFALLAH, N. D. READ and E. U. SELKER, 2004 HPI is essential for DNA methylation in *Neurospora*. *Mol. Cell* **13**: 427–434.
- GALAGAN, J. E., and E. U. SELKER, 2004 RIP: the evolutionary cost of genome defense. *Trends Genet.* **20**: 417–423.
- GALAGAN, J. E., S. E. CALVO, K. A. BORKOVICH, E. U. SELKER, N. D. READ *et al.*, 2003 The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* **422**: 859–868.
- GOODWIN, S. B., J. R. CAVALETTI, C. WAALWIJK and G. H. KEMA, 2001 DNA fingerprint probe from *Mycosphaerella graminicola* identifies an active transposable element. *Phytopathology* **91**: 1181–1188.
- IHGSC, 2001 International Human Genome Sequencing Consortium. *Nature* **409**: 860–921.
- KOUZMINOVA, E., and E. U. SELKER, 2001 dim-2 encodes a DNA methyltransferase responsible for all known cytosine methylation in *Neurospora*. *EMBO J.* **20**: 4309–4323.
- LINARDOPOULOU, E. V., E. M. WILLIAMS, Y. FAN, C. FRIEDMAN, J. M. YOUNG *et al.*, 2005 Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. *Nature* **437**: 94–100.
- LITI, G., and E. J. LOUIS, 2005 Yeast evolution and comparative genomics. *Annu. Rev. Microbiol.* **59**: 135–153.
- LOUIS, E. J., E. S. NAUMOVA, A. LEE, G. NAUMOV and J. E. HABER, 1994 The chromosome end in yeast: its mosaic nature and influence on recombinational dynamics. *Genetics* **136**: 789–802.
- LUNDBLAD, V., and E. H. BLACKBURN, 1993 An alternative pathway for yeast telomere maintenance rescues est1- senescence. *Cell* **73**: 347–360.
- MARTIN, F., A. AERTS, D. AHREN, A. BRUN, E. G. DANCHIN *et al.*, 2008 The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* **452**: 88–92.
- MCCARREY, J. R., and K. THOMAS, 1987 Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene. *Nature* **326**: 501–505.
- MCEACHERN, M. J., and E. H. BLACKBURN, 1996 Cap-prevented recombination between terminal telomeric repeat arrays (telomere CPR) maintains telomeres in *Kluyveromyces fragilis* lacking telomerase. *Genes Dev.* **10**: 1822–1834.
- MCEACHERN, M. J., A. KRAUSKOPF and E. H. BLACKBURN, 2000 Telomeres and their control. *Annu. Rev. Genet.* **34**: 331–358.
- MIURA, A., S. YONEBAYASHI, K. WATANABE, T. TOYAMA, H. SHIMADA *et al.*, 2001 Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis*. *Nature* **411**: 212–214.
- PARDUE, M. L., and P. G. DEBARYSHE, 2003 Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres. *Annu. Rev. Genet.* **37**: 485–511.
- PERKINS, D. D., M. FREITAG, V. C. POLLARD, L. A. BAILEY-SHRODE, E. U. SELKER *et al.*, 2007 Recurrent locus-specific mutation resulting from a cryptic ectopic insertion in *Neurospora*. *Genetics* **175**: 527–544.
- RABINOWICZ, P. D., K. SCHUTZ, N. DEDHIA, C. YORDAN, L. D. PARNELL *et al.*, 1999 Differential methylation of genes and retrotransposons facilitates shotgun sequencing of the maize genome. *Nat. Genet.* **23**: 305–308.
- RONG, Y. S., and K. G. GOLIC, 2003 The homologous chromosome is an effective template for the repair of mitotic DNA double-strand breaks in *Drosophila*. *Genetics* **165**: 1831–1842.
- SAITOU, N., and M. NEI, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406–425.
- SAMBROOK, J., and D. W. RUSSELL, 1989 *Molecular Cloning*. Cold Spring Harbor Laboratory Press, Plainview, NY.
- SELKER, E. U., E. B. CAMBARERI, B. C. JENSEN and K. R. HAACK, 1987 Rearrangement of duplicated DNA in specialized cells of *Neurospora*. *Cell* **51**: 741–752.

- SINGER, T., C. YORDAN and R. A. MARTIENSSSEN, 2001 Robertson's Mutator transposons in *A. thaliana* are regulated by the chromatin-remodeling gene *Decrease in DNA Methylation (DDM1)*. *Genes Dev.* **15**: 591–602.
- SONG, L., S. R. JAMES, L. KAZIM and A. R. KARP, 2005 Specific method for the determination of genomic DNA methylation by liquid chromatography-electrospray ionization tandem mass spectrometry. *Anal. Chem.* **77**: 504–510.
- STUKENBROCK, E. H., S. BANKE, M. JAVAN-NIKKHAH and B. A. McDONALD, 2007 Origin and domestication of the fungal wheat pathogen *Mycosphaerella graminicola* via sympatric speciation. *Mol. Biol. Evol.* **24**: 398–411.
- TAMARU, H., and E. U. SELKER, 2001 A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. *Nature* **414**: 277–283.
- THOMPSON, J. D., T. J. GIBSON, F. PLEWNIK, F. JEANMOUGIN and D. G. HIGGINS, 1997 The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**: 4876–4882.
- TRASK, B. J., C. FRIEDMAN, A. MARTIN-GALLARDO, L. ROWEN, C. AKINBAMI *et al.*, 1998 Members of the olfactory receptor gene family are contained in large blocks of DNA duplicated polymorphically near the ends of human chromosomes. *Hum. Mol. Genet.* **7**: 13–26.
- WATTERS, M. K., T. A. RANDALL, B. S. MARGOLIN, E. U. SELKER and D. R. STADLER, 1999 Action of repeat-induced point mutation on both strands of a duplex and on tandem duplications of various sizes in *Neurospora*. *Genetics* **153**: 705–714.
- WESSLER, S. R., 2006 Eukaryotic transposable elements: teaching old genomes new tricks, pp. 138–162 in *The Implicit Genome*, edited by L. H. CAPOREALE. Oxford University Press, New York.
- XU, J. R., Y. L. PENG, M. B. DICKMAN and A. SHARON, 2006 The dawn of fungal pathogen genomics. *Annu. Rev. Phytopathol.* **44**: 337–366.

Communicating editor: E. U. SELKER

GENETICS

Supporting Information

<http://www.genetics.org/cgi/content/full/genetics.110.117408/DC1>

Accidental Amplification and Inactivation of a Methyltransferase Gene Eliminates Cytosine Methylation in *Mycosphaerella graminicola*

Braham Dhillon, Jessica R. Cavaletto, Karl V. Wood and Stephen B. Goodwin

Copyright © 2010 by the Genetics Society of America
DOI: 10.1534/genetics.110.117408

FIGURE S1.—Sequence alignment of 16 *fam1* elements identified by RECON in the genome of *Mycosphaerella graminicola*. The two longest elements, **ele799** and **ele9870** (green and orange boxes, respectively) were used as reference sequences for *fam1*. Location of the **MgDNMT** region in all the elements is indicated by the yellow box.

FIGURE S2.—Similarity of *fam 1* elements to a LTR retrotransposon family (*fam 15*) in the genome of *Mycosphaerella graminicola*. Representative members of each family are shown as horizontal bars. A 220-bp non-coding region shared between the LTR retrotransposon and *fam 1* elements is shown as a red box. The blue shaded connecting regions are the areas of similarity between these elements.

MgDNMT_deRIP/160-1440 S1_Dim2/1-1281 fam1_01/1-1281 fam1_02/1-1281 fam1_03/1-1282 fam1_04/1-1282 fam1_05/1-582 fam1_06/1-156 fam1_08/1-1281 fam1_09/1-1281 fam1_10/1-1281 fam1_11/1-1281 fam1_12/1-1281 fam1_13/1-697 fam1_14/1-1281 fam1_15/1-1256 fam1_16/1-1281 fam1_17/1-651 fam1_18/1-1281 fam1_19_20/1-1281 fam1_21/1-607 fam1_22/1-553 fam1_23/1-651 fam1_24/1-536 fam1_25/1-349

MKRRISGDTTIVIKDDEDNIVVGPTSRPSKKPRLQNGQLQMDILRAGASLSNAIVISPAAS
MKRRISGDTTIVIEDDEDNIVVGSASRPSKKPRLQNGQLRLDILRAGASQSTAIVVSPAAS
IKRRISGDTTIVIKDNKDNIIVGPTSRPSKKPRLXNRQLXIDILRARASLSNAIISPTAS
IKRRISGDTTIIIXDNKDNIIVGPTSRPSKKPRLXNRQLQIDILRARASLSNAIVISPTAS
IKRRISGDTTIVIKDDEDNIIIVGPTSRPSKKPRLXNGQLXIDILRARASLSNAIVISPTAS
IKRRISGDTTIVIKDDEDNIVVGPTSRPSKKPRLXNGQLXIDILRARASLSNAIVISPAAS
MKRRISGDTTIVIQDDEDNIVVGPTSRPSKKPRLQNGQLXIDILRTRASLSNAIVISPAAS

IKRRISGDTTIVIXDDEDNIIIVGPTSKPSKKPRLXNRQLQIDILRARASLSNAIISPTAS
IKRRISGDTTIVIXDDEDNIIIVGPTSKPSKKPRLXNRQLQIDILRARASLSNAIISPTAS
IKRRISGDTTIVIKDDEDNIIIVGPTSRPSKKPRLQNGQLXIDILRARASLSNAIVISPTAS
IKRRISGDTTIIIXDNEDNIIIVGPTSRPSKKPRLQNGQLXMDILRARASLSNAIVISPAAS
MKRRISGDTTIVIEDDEDNIVVGPTSRPSKKPRLQNGQLXMDILRARASLSNAIVISPAAS
MKRRISGDTTIVIKDDEDNIVVGPTSRPSKKPRLQNGQLQMDILRAGASLSNAIVISPAAS
MKRRISGDTTIVIKDNEDNIIIVGPTSRPSKKPRLXNGQLXIDILRARASLSNAIVISPTAS
IKRRISGDTTIVIXDDKDNIIVGPTSRPSKKPRLQNGQLXIDILRARASLSNAIVISPTAS
MKRRISVDTTIVIKDDEDNIVVGPTSRPSKKPRLRNGQLQMDILRARASLSNAIVISPAAS
MKRQISSDTTIVIKDDEDNIVVGPTSRPSKKPRLXNGQLQIDILRARASLSNAIVISPAAS
IKRRISGDTTIVIKDNEDNIIIVGPTSRPSKKPRLXNRQLXIDILRARASLSNAIISPTAS
IKRRISGDTTIVIXDNEDNIIIVGPTSRPSKKPRLXNRQLXIDILRARASLSNAIVISPTAS
IKRRISGDTTIVIKDDEDNIIIVGPTSRPSKKPRLQNRQLQIDILRAGASLSNAIVISPTAS
MKRRISVDTTIVIKDDEDNIVVGPTSRPSKKPRLRNGQLQMDILRARASLSNAIVISPAAS

IKRRISGDTTIVIQDDEDNIIIVGPTSRPSKKPRLXNRQLQIDILRARASLSNTIVISPTAS

MgDNMT_deRIP/160-1440 S1_Dim2/1-1281 fam1_01/1-1281 fam1_02/1-1281 fam1_03/1-1282 fam1_04/1-1282 fam1_05/1-582 fam1_06/1-156 fam1_08/1-1281 fam1_09/1-1281 fam1_10/1-1281 fam1_11/1-1281 fam1_12/1-1281 fam1_13/1-697 fam1_14/1-1281 fam1_15/1-1256 fam1_16/1-1281 fam1_17/1-651 fam1_18/1-1281 fam1_19_20/1-1281 fam1_21/1-607 fam1_22/1-553 fam1_23/1-651 fam1_24/1-536 fam1_25/1-349

EQDRGESKQEEEDGAQRAVQYKMESGEEQTSQTVGDREVEDQGLLRTPRQRRRLRTALTTP
ERQGEGSQEEAGAAAVQYEVESGEEQTSQTIIGQQVADQGTTRTPRQRRFGTALTTP
KXDRGESKQEEEDRVQRVXYKIESGEEEXTSXTVGDREVVDXGLLRTPRQRRRLRTLTPL
KXDRGESKXXEEDRVXRAVXYKIESGEEQTSXTVGNXEVKDXGLLRTPRQRRRLRTALTPL
KQDRGELKQXEEEDRVXRAVXYKIESGEEEXTSXTVGDREVVDXGLLRTPRXRRRLRTLTPL
KQDRGELKQXEEEDRVXRAVXYKIESGEEEXTSXTVGDREVVDXGLLRTPRXRRRLRTLTPL
EQDRGESKQEEEDGVQRAVXYKIELGEEQTSQTVGDREVVDXGPLRTPRQRRRLRTALTTP

KQDRGESKQEEEDRVQRAVXYKIESGEEQTSQTVGDREVVDXGLLRTPRQRRRLRTALTTP
KQDRGESKQEEEDRVQRAVXYKIESGEEQTSQTVGDREVVDXGLLRTPRQRRRLRTALTTP
KQDRGESKQEEEDRVQRAVXYKIESGEEQTSQTVGDREVVDXGLLRTPRXRRRLRTILTPL
KQDXGESKQEEEDRVQRAVXYKIESGEEQTSXTVGDREVVDXGLLRTPRQRRRLRTLTPL
KQDRGESKQEEEDGVQRAVXYKIESGEEQTSQTVGDREVEDQGLLRTPRQRRRLRTLTTP
EQDRGESKQXEEEDGVQRAVXYKIESGEEEXTSXTVGDREVEDQGLLRTPRQRRRLRTALTPL
KQDRGESKQEEEDGVQRAVXYKIESGEEQTSQIVGDREVVDXGLLRTPRQRRRLRTILTPL
EXDXGESKXXEEDRVXRAVXYKIELGEEQTSQTVGDREVVDXGLLRTPRQRRRLRTVLTPL
KQDRGESKQEEEDRVQRAVXYKIESGEEQTSQTVGDREVVDXGLLRTPRQRRRLRTLTPL
EQDRGESKQEEEDGVQRAVQYKMESGEEQTSQTVGDREVEDQGPLRTPRQRRRLRTLTTP
KXDRGESKQEEEDGVQRAVXYKIESGEEQTSXTVGDREVVDXGLLRTPRQRRRLRTALTPL
KQDRGESKXXEEDRAXRAVXYKIELGEEQTSXTVGDREVVDXGLLRTPKXRRRLRTILTPL
KXDRGESKQXEEEDRVXRAVXYKIESGEEQTSXTVGDREVVDXGLLRTPRQRRRLRTILTPL
KXDRGESKQEEEDGVQRAVXYKIESGEEQTSQTVGDREVVDXGLLRTPRQRRRLRTALTTP
EQDRGESKQEEEDGVQRAVQYKMESGEEQTSQTVGDREVEDQGPLRTPRQRRRLRTLTTP

KXDRGESKQEEEDRVQRAVQYKIELGEEQTSXTVGDREVVDXGLLRTPRQRRRLRTLTTP

MgDNMT_deRIP/160-1440 S1_Dim2/1-1281 fam1_01/1-1281 fam1_02/1-1281 fam1_03/1-1282 fam1_04/1-1282 fam1_05/1-582 fam1_06/1-156 fam1_08/1-1281 fam1_09/1-1281 fam1_10/1-1281 fam1_11/1-1281 fam1_12/1-1281 fam1_13/1-697 fam1_14/1-1281 fam1_15/1-1256 fam1_16/1-1281 fam1_17/1-651

ATETELGDEPAARAEPGATARSKKRQLRYLSHIS IAGPAFPRSAYHGYKPPLQPVNEVE
ATETKSEDEPAARAEPGTARSKKRQLRYLSHISNAGPAFPRSAYHGYIPPLQPVNEVE
ATETELRDEPTTARAEPRAIARSGKRQLRYLSYISIIIGLAFPRSAYYGYKPPLXPINEVE
ATEIELGDEPATRAEPRAIARSGKRXXLRYLSYISIIIGLAFPRSAYHGYKPPLQPINKEVE
ATETELGDEPTARAEPGAIAARSGKKRQLCYLSYISIIIGLAFPRSAYYGYKPPLQPVNEVE
ATETELGDEPATRAEPRAIARSGKKRQLRYLSYISIIIGLAFPRSAYYGYKPPLQPVNEVE
ATETELGDEPATRAEPGATARSKKRQLRYLSHISIVGPAFPRSAYYGYKPPLQPVNEVE

ATETELGDEPATRAEPGATARSKKRQLRYLSHIS IAGLAFPRSAYYGYKPPLQPMNEVE
ATETELGDEPATRAEPGATARSKKRQLRYLSHIS IAGLAFPRSAYYGYKPPLQINEVE
ATETELGDEPTARAEPGAIAARSGKKRQLRYLSHISIVGPAFPRSAYYGYKPPLQPVNEVE
ATETEPGDEPAARAEPGATARSKKRQLRYLSHISIIIGPAFPRSAYHGYKPPLQPVNEVE
ATETELGDEPAARAEPGATARSKKRQLRYLSHIS IAGLAFPRSAYHGYKPPLQINEVE
ATETELGDEPAARAEPGATARSKKRQLRYLSHISIVGPAFPRSAYYGYKPPLQPVNEVE
ATETELGDEPAARAEPGATARSKKRQLCYLSYISIVGPAFPRSAYYGYKPPLQPVNEVE
ATETELGDEPAARAEPGATARSKKRQLRYLSHISIVGPAFPRSAYYGYKPPLQPVNEVE
ATETELGDEPTTARAEPGATARSKKRQLRYLSHISITGPAFPRSAYYGYKPPLQINEVE
ATETEPGDEPAARAEPGATARSKKRQLRYLSHIS IAGPAFPRSAYHGYKPPLQPVNEVE

```
fam1_18/1-1281 ATETELGDEPAARAEPKAIARSGKRQXLRYSYISITGPAFPRSAIYGYKPLQPVNEVE
fam1_19_20/1-1281 ATEIELGDEPTTAEPRAIARSGKRQRLRYSYISIIGLAFPRSAIYGYKPLQPVNEVE
fam1_21/1-607 ATETEPGDEPATRVESTRATARSQKRQRLRYSYISIVGLAFPRSAIYGYKPLQPVNEVE
fam1_22/1-553 ATETELGDEPTTAEPRATARSQKRQRLRYSYISIVGLAFPRSAIYGYKPLQPVNEVE
fam1_23/1-651 ATETEPGDEPAARAEPGATARSQKRQRLRYSYISITGPAFPRSAIYGYKPLQPVNEVE
fam1_24/1-536 -----
fam1_25/1-349 ATETELGDEPTTAEPRATARSQKRQRLRYSYISITGPAFPRSAIYGYKPLQPVNEVE
```

```
MgDNMT_deRIP/160-1440 ASLDLSAGFFAAAKSAGTETPECATFKLDQFSVYIADGRRASDLVTLEKLRQSNYAEFCF
S1_Dim2/1-1281 ASLDLSAGFFAAAKSAGTETPEYATFKLDQFSVYIADGRRASDLVTLEKLRQSNYAEFCF
fam1_01/1-1281 VSLDLSAGFFAAAKSTSTETPECTTFKLDXFSVYIADGRRASDLVTLEKLRQSNYAEFCF
fam1_02/1-1281 ASLDLSAGFFAAAKSTSTETPECTTFKLDQFSVYIADGRRASDLVTLEKLRQSNYAEFCF
fam1_03/1-1282 VSLDLSVGFFAAAKSASTETPECATFKLDQFSVYIADGRRASDLVTLEKLRQSNYAEFCF
fam1_04/1-1282 VSLDLSVGFFAAAKSASTETPECATFKLDQFSVYIADGRRASDLVTLEKLRQSNYAEFCF
fam1_05/1-582 ASLDLSAGFFAAAKSASTETPECATFKLDQFSVYIADGRRASDLVTLEKLRQSNYAEFCF
fam1_06/1-156 -----
fam1_08/1-1281 ASLDLSAGFFAAAKSTSTETPECTTFKLDQFSVYIADGRRASDLVTLEKLRQSNYAEFCF
fam1_09/1-1281 ASLDLSAGFFAAAKSTSTETPECTTFKLDQFSVYIADGRRASDLVTLEKLRQSNYAEFCF
fam1_10/1-1281 ASLDLSAGFFIATKSAGTETPECATFKLDXFLVYIADGRRASDLVTLEKLRQSNYAEFCF
fam1_11/1-1281 ASLDLSVGFFAAAKSASTETPEYATFKLDXFSVYIADGRRASDLVTLEKLRQSNYAEFCF
fam1_12/1-1281 ASLDLSVGFFAAAKSASTETPECATFKLDQFSVYIADGRRASDLVTLEKLRQSNYAEFCF
fam1_13/1-697 ASLDLSAGFFAAAKSAGTETPECATFKLDQFSVYIADGRRASDLVTLEKLRQSNYAEFCF
fam1_14/1-1281 ASLDLSAGFFAAAKSASTETPECATFKLDQFSVYIADGRRASDLVTLEKLRQSNYAEFCF
fam1_15/1-1256 ASLDLSIGFFVAAAKSTSTKTECATFKLDXFSVYIADGRRASDLVTLEKLRQSNYAEFCF
fam1_16/1-1281 ASLDLSAGFFAAAKSASTETPECATFKLDQFSVYIADGRRASDLVTLEKLRQSNYAEFCF
fam1_17/1-651 ASLDLSAGFFAAAKSAGTETPECATFKLDQFSVYIADGRRASDLVTLEKLRQSNYAEFCF
fam1_18/1-1281 VSLDLSVGFFAAAKSASTETPECATFKLDXFSVYIADGRRASDLVTLEKLRQSNYAEFCF
fam1_19_20/1-1281 ASLDLSAGFFAAAKSTSTETPECTTFKLDXFLVYIADGRRASDLVTLEKLRQSNYAEFCF
fam1_21/1-607 ASLDLSVGFFVATKSTSTETPECATFKLDXFLVYIADGRRASDLVTLEKLRQSNYAEFCF
fam1_22/1-553 ASLDLSVGFFAAAKSAGTETPECATFKLDXFSVYIADGRRASDLVTLEKLRQSNYAEFCF
fam1_23/1-651 ASLDLSAGFFAAAKSASTETPECATFKLDQFSVYIADGRRASDLVTLEKLRQSNYAEFCF
fam1_24/1-536 -----
fam1_25/1-349 ASLDLSVGFFAATKSTSTETPECTTFKLDXFLVYIADGRRASDLVTLEKLRQSNYAEFCF
```

```
MgDNMT_deRIP/160-1440 DGILSAGNEKRYVQGVRFELAISRYGDLDFDFAAGSICIXSKLQSKASALFGHWYELGT
S1_Dim2/1-1281 DGILSAGNEKRYVQGVRFELAISRYGDLDFDFAAGSICIXSKLQSKASALFGHWYELGT
fam1_01/1-1281 DSILSVGNEKRYVXGVFEILAISRYGDLDFDFAAGSICIXSKLXSKASALFSYXYELGT
fam1_02/1-1281 DSILLAGNEKRYVXGVRFELAISRYGDLDFDFAAGSICIXSKLQSKASALFSYXYELGT
fam1_03/1-1282 DSILLAGNEKRYVXGIRFEILAISRYGDLDFDFAAGSICIXSKLXSKASALFSYXYELGT
fam1_04/1-1282 DSILLAGNEKRYVXGIRFEILAISRYGDLDFDFAAGSICIXSKLXSKASALFSYXYELGT
fam1_05/1-582 DGILSAGNEKRYVXGVFEILAISRYGDLDFDFAAGSICIXSKLXSKASALFSYXYELGT
fam1_06/1-156 -----
fam1_08/1-1281 DSILSAGNEKRYVQGVRFELAISRYGDLDFDFAAGSICIXSKLQSKASALFSYXYELGT
fam1_09/1-1281 DSILSAGNEKRYVQGVRFELAISRYGDLDFDFAAGSICIXSKLQSKASALFSYXYELGT
fam1_10/1-1281 DGILSAGNEKRYVXGVFEILAISRYGDLDFDFAAGSICIXSKLXSKASALFSYXYELGT
fam1_11/1-1281 DSILSAGNEKRYVXGVFEILAISRYGDLDFDFAAGSICIXSKLQSKASALFSYXYELGT
fam1_12/1-1281 DSILLAGNKKRYVQGVFEILAISRYGDLDFDFAAGSICIXSKLQSKASALFSYXYELGT
fam1_13/1-697 DGILSAGNKKRYVQGVFEILAISRYGDLDFDFAAGSICIXSKLQSKASALFGHWYELGT
fam1_14/1-1281 DSILSVGNEKRYVXGVFEILAISRYGDLDFDFAAGSICIXSKLQSKASALFGHWYELGT
fam1_15/1-1256 DSILSVGNEKRYVXGMRFEILAISRYGDLDFDFAAGSICIXSKLQSKASTLFSYXYELGT
fam1_16/1-1281 DSILSAGNKKRYVXGIRFEILAISRYGDLDFDFAAGSICIXSKLXSKASALFSYXYELGT
fam1_17/1-651 DGILSAGNEKRYVQGVRFELAISRYGDLDFDFAAGSICIXSKLQSKASALFGHWYELGT
fam1_18/1-1281 DGILSVGNKKRYVXGVFEILAISRYGDLDFDFAAGSICIXSKLQSKASALFSYXYELGT
fam1_19_20/1-1281 DSILSAGNKKRYVXGVFEILAISRYGDLDFDFAAGSICIXSKLQSKASALFSYXYELGT
fam1_21/1-607 DSILSAGNEKRYVXGVFEILAISRYGDLDFDFAAGSICIXSKLQSKASALFSYXYELGT
fam1_22/1-553 DSILSAGNEKRYVXGVFEILAISRYGDLDFDFAAGSICIXSKLQSKASALFGHWYELGT
fam1_23/1-651 DGILSAGNEKRYVQGVRFELAISRYGDLDFDFAAGSICIXSKLQSKASALFGHWYELGT
fam1_24/1-536 -----
fam1_25/1-349 DSILSAGNEKRYVQGVFEILAISRYGDLDFDFAAGSICIXSKLQSKASALFSYXYELGT
```

```
MgDNMT_deRIP/160-1440 PAPEYRRFYEPFLWLARFGKHFDVYLLSNXQVTLGHFRRRFYQWLTGLYGDNTSFNXXLA
S1_Dim2/1-1281 PAPEYRRFYEPFLWLARFGKHFDVYLLSTQVTLGHFRRRFYQWLTGLYGDNTSFNXXLA
fam1_01/1-1281 PAPEYRRFYEPFLXLVQFGKHFDVYLLSNXQVTLRHFRFRFYXXLTRLYSDNTSFNXXLA
fam1_02/1-1281 PILEYRRFYELFLXLVRFGKHFDVYLLSNXQVTLRHFRFRFYXXLTRLYSDNTSFNXXLA
fam1_03/1-1282 PTPEYRRFYEPFLXLARFGKHFDVYLLSNXQVTLRHFRFRFYXXLTRLYSDNTSFNXXLA
fam1_04/1-1282 PIPEYRRFYEPFLXLARFGKHFDVYLLSNXQVTLGHFRRRFYXXLTRLYSDNTSFNXXLA
fam1_05/1-582 PAPEYRRFYEPFLXLQFGKHFDVYLLSNQVTLGHFRRRFYQWLTGLYGDNTSFNXXLA
fam1_06/1-156 -----
fam1_08/1-1281 PTPEYRRFYEPFLXLQFGKHFDVYLLSNXQVTLRHFRFRFYXXLTRLYSDNTSFNXXLA
fam1_09/1-1281 PTLEYRRFYEPFLXLQFGKHFDVYLLSNXQVTLRHFRFRFYXXLTRLYSDNTSFNXXLA
fam1_10/1-1281 PAPEYRRFYEPFLWLVRFGKHFDVYLLSNXQVTLGHFRRRFYXXLTRLYSDNTSFNXXLA
fam1_11/1-1281 PTPEYRRFYEPFLXLVRFGKHFDVYLLSNXQVTLGYFRFRFYXXLTRLYSDNTSFNXXLA
fam1_12/1-1281 PAPEYRRFYELFLXLARFGKHFDVYLLSNQVTLGHFRRRFYQWLTGLYGDNTGFNXXLA
```

```

fam1_13/1-697      PAPEYRRFYEPFLWLARFGKHFDVYLLSNXQITLGHFRRRFYXWLTGLYGDDTSFNXXWLA
fam1_14/1-1281     PAPEYCRFYEPFLXLAQFGKHFDVYLLSNXQVTLRHFRRRFYXXLTGLYGDNSTFNXXLA
fam1_15/1-1256     PTPEYRRFYEPFLXLAQFGKHFDVYLLSNXQVTLRHFRRRFYQXLTRLYGDNSTFNXXLA
fam1_16/1-1281     PTPEYRRFYEPFLXLAQFGKHFDVYLLSNXQVTLGHFRRRFYXWLTGLYGDNSTFNXXLA
fam1_17/1-651      PAPEYRRFYEPFLWLARFGKHFDVYLLSNQVTLGHFRRRFYQWLTGLYGDDTGFNQWLA
fam1_18/1-1281     PALEYRRFYEPFLXLAQFGKHFDVYLLSNXQITLRFHRRRFYXXLTGLYGDNSTFNXXLA
fam1_19_20/1-1281 PTPEYCRFYEPFLXLVXFGKHFDVYLLSNXXVTLRHFRRRFYXXLTRLYSDNSTFNXXLA
fam1_21/1-607      PIPEYRRFYKPFLLXLAQFGKHFDVYLLSNXQVTLRHFRRRFYXXLTGLYSDNSTFNXXLA
fam1_22/1-553      PAPEYRRFYKPFLLXLAQFGKHFDVYLLSNXQVTLRHFRRRFYQXLTRLYGDDTSFNQXLA
fam1_23/1-651      PAPEYRRFYEPFLWLARFGKHFDVYLLSNQVTLGHFRRRFYQWLTGLYGDDTGFNQWLA
fam1_24/1-536      -----
fam1_25/1-349      PAPEYRRFYELFLXLAQFGKHFDVYLLSNXQVTLPIYIXLGYTRXRQRF-----

```

```

MgDNMT_deRIP/160-1440 ECKLKDFRTTIAANVGFLWKECWDIDDKESGLCRQPIWGEVDPYNLTAILAQPKNELRTI
S1_Dim2/1-1281      KSKLEDFRTTIAANVGFLWKECWDIDDKESGLCRHPIWGEVDPYNLKAIPAQSNKELRTI
fam1_01/1-1281      KCKLKDFRTTIAANVGFLXKECXDIDNKESRLYXQPIXGKVDPNYNTAILAXLNKELRTI
fam1_02/1-1281      KYKLDKDFRTTIAANVGFLWKECXDIDNKESRLCRQPIXGKVDPNYNTAILAXPNKELRTI
fam1_03/1-1282      ECKLKDFRTTIAANVGFLXKECXDIDNKESGLCXQPIXGKVDPNYNTAILAXLNKELRTI
fam1_04/1-1282      ECKLKDFRTTIAANVGFLXKECXDIDNKESGLCXQPIXGKVDPNYNTAILAXLNKELRTI
fam1_05/1-582       ECKLKDFRTTIAANVGFLWKECWDIDDKESGLCRQPIXGEVDPYNLTAILAQPKNELRTI
fam1_06/1-156       -----
fam1_08/1-1281      KCKLKDFRTTIAAKVGFLXKECXDINNKESSLRQPIXGKVNPNYNTAILAQPKNELRTI
fam1_09/1-1281      KCKLKDFRTTIAAKVGFLXKECXDINNKESSLRQPIXGKVNPNYNTAILAQPKNELRTI
fam1_10/1-1281      EYKLDKDFRTTIAANVGFLXKECXDIDNKESRLCRQPIXGKVDPNYNTAILAQPKNELRTI
fam1_11/1-1281      ECKLKDFRTTIAANVGFLWKECXDINNKESSLRQPIXGEVDPYNLTAILAQPKNELRTI
fam1_12/1-1281      ECKLKDFRTTIAANVGFLWKECWDIDDKESRLCXQPIXGKVDPNYNTAILAQPKNELRTI
fam1_13/1-697       ECKLKDFYTTIAANVGFLWKECWDINNKESSLRQPIWGEVDPYNLTAILAQPKNELRTI
fam1_14/1-1281      ECKLKDFRTTIAANVGFLWKECWDIDNKESGLCRQPIXGKVDPNYNTAILAQPKNELRTI
fam1_15/1-1256      KCKLKDFRTTIAANVGFLXKECXDIDNKESRLCRQPIXGKVDPNYNTAILAXLNKELQTI
fam1_16/1-1281      ECKLKDFRTTIAANVGFLXKECXDIDDKESRLCRQPIXGKVDPNYNTAILAXPNKELRTI
fam1_17/1-651       ECKLKDFRTTIAANVGFLWKECWDIDDKESGLCRQPIWGEVDPYNLTAILAQPKNELRTI
fam1_18/1-1281      KCKLKDFRTTIAANVGFLWKECXDIDDKESGLCRQPIWGEVDPYNLTAILAQLNKELRTI
fam1_19_20/1-1281 KCKLKDFRTTIAANVGFLXKECXDIDNKESRLCXXPIXGKVDPNYNTAILAQLNKELRTI
fam1_21/1-607       KCKLKDFRTTITNNGFLXKECXDIDNKESKLCRLIXGKVDPNYNTAILVXLNKELRTI
fam1_22/1-553       ECKLKDFRTTIAANVGFLXKECWDIDNKESGLCXQPIWGEVDPYNLTAILAQPKNELRTI
fam1_23/1-651       ECKLKDFRTTIAANVGFLWKECWDIDDKESGLCRQPIWGEVDPYNLTAILAQPKNELRTI
fam1_24/1-536       -----
fam1_25/1-349      -----

```

```

MgDNMT_deRIP/160-1440 VTPFVYQCFRHMFGDQLEERTITDRNVLKKSDCLKARLRLTPRRASPSAPPMQTLTPRS
S1_Dim2/1-1281      VTPFVYQCFRHMFGDQLEERTITDRNVLKKSDCLKARLRLTPRRASPSAPPVQTLTPSS
fam1_01/1-1281      VTPFIYXCFRHIYFGNXLEERTITDRNVLKKSDCLKARLRLTPRRASPSVPLMXTLTPRS
fam1_02/1-1281      VTPFVYXCFXHIYFGNXLEERTITDRNVLKKSDCLKARLRLTLRRASPSVLLMQTLTPRS
fam1_03/1-1282      VTPFVYXCFXHMFGNXLEERTITDRNVLKKSDCLKARLRLTLRRASPSAPLIQTLTPRS
fam1_04/1-1282      VTPFVYXCFXHMFGNXLEERTITDRNVLKKSDCLKARLRLTLRRASPSPTPMQTLTPRS
fam1_05/1-582       VTPFVYQCFRHMFGDQLEERTITDRNVLKKSDCLKARLRLTPRRASPSAPPMQTLTPRS
fam1_06/1-156       -----
fam1_08/1-1281      VTPFVYXCFRHMFGNXLEERTITDRNVLKKSDCLKARLRLTPRRASPSAPPIQTLMPRS
fam1_09/1-1281      VMPFVYXCFRHMFGNXLEERTITDRNVLKKSDCLKARLRLTPRRASPSAPPIQTLMPRS
fam1_10/1-1281      VTPFVYQCFRHMFGNXLEERTITDRNVLKKSDCLKARLRLTPRRASPSALPMQTLTPRS
fam1_11/1-1281      VTPFVYXYFRHMFGNXLEEQITIDRNVLKKSDCLKARLRLTPRRASPSAPLMXTLTPRS
fam1_12/1-1281      VTPFVYQCFRHMFGDQLEERTITDRNVLKKSDCLKARLRLTPRRASLSAPPMQTLTPRS
fam1_13/1-697       VTPFVYQCFRHMFGDQLEERTITDRNVLKKSDCLKARLRLTPRRASPSAPPMQTLTPRS
fam1_14/1-1281      VTPFVYQCFRHMFGDQLEERTITDRNVLKKSDCLKARLRLTPRRASPSALPMQTLTPRS
fam1_15/1-1256      VTPFVYXCFXHMFGNXLEEQITIDRNVLKKSDCLKARLRLTPRRASPSALLMQTLTPRS
fam1_16/1-1281      VTPFVYXCFXHIYFGDQLEERTITDRNVLKKSDCLKARLRLTPRRASPSVPPMQTLTPRS
fam1_17/1-651       VTPFVYQCFRHMFGDQLEERTITDRNVLKKSDCLKARLRLTPRRASPSAPPMQTLTPRS
fam1_18/1-1281      VTPFVYXCFRHMFGDQLEERTITDRNVLKKSDCLKARLRLTPRRASPSALPMQTLTPRS
fam1_19_20/1-1281 VTLFVYXCFRHMFGNXLEERTITDRNVLKKSDCLKARLRLTPRRASPSAPLIQTLTPRS
fam1_21/1-607       VTPFVYXYFRHMFGNXLEERTITDRNVLKKSDCLKVRLRLTPRRASPSALLMQTLTPRS
fam1_22/1-553       VTPFVYQCFRHMFGDQLEERTITDQNVLKKSDCLKVRLRLTPRRASPSAPPMQTLTPRS
fam1_23/1-651       VTPFVYECFRHMFGDQLEERTITDRNVLKKSDCLKARLRLTPRRASPSAPPMQTLTPRS
fam1_24/1-536       -----
fam1_25/1-349      -----

```

```

MgDNMT_deRIP/160-1440 MSRSSEQEGNKAVTNTGDVVCVAPDVSTKWGIKAHAWYAYVQNVHTRKDRSLDVLWLY
S1_Dim2/1-1281      MSRSSEQEWNKAVTNTGDVVCVAPDVSTKWGIKAHAWYAYVQNVHTRKDRSLDVLWLY
fam1_01/1-1281      ISRSSEQKGNKVVTNIGDVVYVAPDVSTKXGIKAYIXYAYVXNVYTRKDRSLDVLXLY
fam1_02/1-1281      MSXSESXEGNKVVTNIGDVVYVAPDISTKXGIKAYTXAYVXNMVYTRKDRSLDVLXLY
fam1_03/1-1282      ISRSSESXEGNKAVTNTGDVVCVAPDVSTKXGIKAHAXYAYVXNVYTRKDRSLDVLXLY
fam1_04/1-1282      ISRSSESXEGNKAVTNTGDVVCVAPDVSTKXGIKAHAXYAYVXNVYTRKDRSLDVLXLY
fam1_05/1-582       MSRSSEQEGNKAVTNTGDVVCVAPDVSTKXGIKAHAWYAYVXNVHTRKDCSLDVLWLY
fam1_06/1-156       -----

```



```

fam1_08/1-1281 MSRSEPQEGNKAVTVNTGDVVYVAPDISTKXGIIKAHAXYAYVQNVHTRKDRSTSLDVLXLY
fam1_09/1-1281 MSRSEPQEGNKAVTVNTGDVVYVAPDISTKXGIIKAHAXYAYVQNVHTRKDRSTSLDVLXLY
fam1_10/1-1281 ISRSESXEGNKAVIVNTGDIVCVAPDVSTKXGIIKAHAXYAYVXNVHTCKDRSTSLDILXLY
fam1_11/1-1281 MSRSESQEGNKVVTVNTGDVVYVAPDVSTKXGIIKAYAXYAYVXNVYTRKDRSTSLDVLXLY
fam1_12/1-1281 MSRSESQEGNKAVMVNTGDVVCVAPDVSTKWGIKAHAWYAYVXNVHTCKDRSTSLDVLWLY
fam1_13/1-697 MSRSESQEGNKAVTVNTGDVVCVAPDVSTKWGIKAHAWYAYVQNVHTRKDRSTSLDVLWLY
fam1_14/1-1281 MSRSESQEGNKAVIVNTGDIVCVAPDVSTKXGIIKAYAXYAYVXNVYTRKDRSTSLDVLXLY
fam1_15/1-1256 ISRSESQEGNKVVTVNTGDIVCVAPDVSTKXGIIKAYXYTVQNMHTRKDRSTSLDVLXLY
fam1_16/1-1281 ISRSESXEGNKAVTVNIGDVVCVAPDVSTKXGIIKAHAXYAYVXNVHTRKDRSTSLDVLXLY
fam1_17/1-651 MSRSESQEGNKAVMVNTGDVVCVAPDVSTKWGIKAHAWYAYVQNVHTCKDRSTSLDVLWLY
fam1_18/1-1281 MSRSESXEGNKAVTVNIGDVVCVAPDVSTKXGIIKAYXYAYVQNVYTRKDRSTSLDVLWLY
fam1_19_20/1-1281 ISXKSXEGNKAVIVNTGDVVCVAPDISTKXGIIKAHAXYAYIXNVHTRKDRSTSLDVLXLY
fam1_21/1-607 ISRSESXEGNKAVTVNIGDVVCVAPDVSTKXGIIKAYAXYAYVQNVHTRKDRSTSLDVLWLY
fam1_22/1-553 MSRSESXEGNKAVTVNTGDVVCVAPDVSTKXGIIKAHAXYAYVXNVHIREDRSTSLDVLXLY
fam1_23/1-651 MSRSESQEGNKAVMVNTGDVVCVAPDVSTKWGIKAHAWYAYVQNVHTCKDRSTSLDVLWLY
fam1_24/1-536 -----
fam1_25/1-349 -----

```

```

MgDNMT_deRIP/160-1440 EPKDTTLGKAYYPFANELFLSDNCSCGTSALDLDCLDKADVSWYAKDPYQGGGHFVRQK
S1_Dim2/1-1281 EPRDTTLGKAYYPFANELFLSDNCSCGTAALDLDCLDKADVSWYAKDPYQGGGHFVRQK
fam1_01/1-1281 ELKDTILSKAYYPFANKLFLSDNYSYISALDLDYVLDKADVXCXYAKDLYXGGGHFVRXK
fam1_02/1-1281 KLKDTILSKAYCPFANKLFLSDNCSYGTSAALDLDCLDKADICXYTKDLYXQGGGHFVRXK
fam1_03/1-1282 ELKDTTLGKAYYPFANKLFLSDNCSYGTSAALDLDYVLDKAEVCXYTKDLYXQGGGHFVRQK
fam1_04/1-1282 ELKDTTLGKAYYPFANKLFLSDNCSYGTSAALDLDYVLDKAEVCXYAKDLYXQGGGHFVRQK
fam1_05/1-582 EPKDTTLGKAYYPFANELFLSDNCSCGTSALDLDCLDKADV-----
fam1_06/1-156 -----KADVXCXYAKDPYXQGGGHFVRQK
fam1_08/1-1281 ELKDTTLGKAYYPFANKLFLSDNYSYGTSAALDLDCLDKADVXCXYTKDLYXQGGGHFVRQK
fam1_09/1-1281 ELKDTTLGKAYYPFANKLFLSDNYSYGTSAALDLDCLDKADVXCXYTKDLYXQGGGHFVRQK
fam1_10/1-1281 ELKDTTLGKAYYPFANKLFLSDNYSYSTSAALDLDCLDKADVXCXYAKDPYQGGGHFVRQK
fam1_11/1-1281 KLKDTTLGKAYYPFANKLFLSDNCSYGTSAALDLDCLDKADVXCXYAKDLYXGGGHFVRQK
fam1_12/1-1281 KPKDTTLGKAYYPFANELFLSDNCSCGTLALDLDCLDKVDVCWYAKDPYXQGGGHFVRQK
fam1_13/1-697 EPKDTTLGKAYYPFANELFLSDNCSCGTSALDLDCLDKADVVCWYAKDPYQGGGHFVRQK
fam1_14/1-1281 KLKDTTLGKAYYPFANELFLSDNCSYGTSAALDLDCLDKADVYXKADLYXQGGGHFVRQK
fam1_15/1-1256 ELKDTTLGKAYYPFANKLFLSDNCSYGTSAALDLDYVLDKADVXCXYTKDLYXGGGHFVRQK
fam1_16/1-1281 ELKDTTLGKAYYPFANKLFLSDNCSYGTSAALDLDCLDKADVXCXYAKDLYXQGGGHFVRQK
fam1_17/1-651 EPKDTTLGKAYYPFANELFLSDNCSCGTSALDLDCLDKADVVCWYAKDPYQGGGHFVRQK
fam1_18/1-1281 ELKDTTLGKAYYPFANELFLSDNYSYGTSAALDLDCLDKADVXYXKADLYXQGGGHFVRQK
fam1_19_20/1-1281 ELKDTTLGKAYYPFANKLFLSDNCSYGTSAALDLDCLDKADVXYTKDLYXQGGGHFVRQK
fam1_21/1-607 KLKDTILSNYYPFANKLFLSDNCSYGTSAALDLDCLDKADVXCXYAKDLYXQGGGHFVRQK
fam1_22/1-553 ELKDTTLGKAYY-----
fam1_23/1-651 ELKDTTLGKAYYPFANELFLSDNCSCGTSALDLDCLDKADVVCWYAKDPYQGGGHFVRQK
fam1_24/1-536 -----
fam1_25/1-349 -----

```

```

MgDNMT_deRIP/160-1440 FRTVHEQDITYDFVAIQASDFRCQCKHDVPIFDECRSKYAI GDTVLVREWNDDLSEDRLOP
S1_Dim2/1-1281 FRTVHEQDITYDFVAIQASDFRCQCKDDVPIFDECRSKYAI GDTVLVREWNDDLSEDRLOP
fam1_01/1-1281 FYIVYEQDIYDFVVI XASDFRCXCKHNVPIFDECRSKYAI SNTVLRVREXNDLYEDRLXL
fam1_02/1-1281 FCIVYEXDITYDFVTI QASDFRCXCKHNVPIFDECRSKYAI SNTVLRVREWNNDLNEDEXLXL
fam1_03/1-1282 FRTVHEQDITYDFVAI XASDFRCXCKHNVPIFDECRSKYAI SNTVLRVREXNDLNEDEXLXL
fam1_04/1-1282 FRTVHEQDITYDFVAI XASDFRCXCKHNVPIFDECRSKYAI SNTVLRVREXNDLNEDEXLXL
fam1_05/1-582 -----
fam1_06/1-156 FRTVYEQDITYDFVAIQASDFRCQCKHDVPIFDECRSKYAI GDTVLVREWNDDLSEDRLOP
fam1_08/1-1281 FRTVHEQDIYDFVAI XASDFRCXCKHNVPIFDECRSKYAI SNTVLRVREXNDLNEDEXLXL
fam1_09/1-1281 FRTVHEQDIYDFVAI XASDFRCXCKHNVPIFDECRSKYAI SNTVLRVREXNDLNEDEXLXL
fam1_10/1-1281 FRIVHEQDITYDFVVI QVSDFRXCXCKYVPIFDECRSKYAI SNTVLRVREWNNDLNEDEXLXL
fam1_11/1-1281 FRTVYEQDIYDFVAI XASDFRCXCKHNVPIFDECRSKYAI SNTVLRVREXNDLNEDEXLXL
fam1_12/1-1281 FRTVHEQDITYDFVAIQVSDFRYCQCKHNVPIFDECRSKYAI GDTVLVREWNDDLSEDRLOP
fam1_13/1-697 FRTVHEQDITYDFVAIQASDFRCQCKHDVPIFDECRSKYAI GDTVLVREWNDDLSEDRLOP
fam1_14/1-1281 FRIVHEQDITYDFVAIQASDFRCQCKHNVPIFDECRSKYAI GDTVLVREXNDLNEDEXLXL
fam1_15/1-1256 FRTVYEQDITYDFVTIXVSDFRXCXCKYVPIFDECRSKYAI SNTVLRVREXNDLNEDEXLXL
fam1_16/1-1281 FRTVHEXDTYDFVAIQVSDFRXCXCKHDVPIFDECRSKYAI GDTVLVREXNDLNEDEXLXL
fam1_17/1-651 FRTVHEQDITYDFVAIQASDFRCQCKHDVPIFDECRSKYAI GDTVLVREWN-----
fam1_18/1-1281 FRIVYEQDIYDFVVI XASDFRCQYKHDVPIFDECRSKYI SNTVLRVREWNNDLSEDRLOP
fam1_19_20/1-1281 FRIVYEQDIYDFVAI XASDFRCXCKHNVPIFNECRSKYI SNTVLRVREXNDLSEDRLOP
fam1_21/1-607 SGEYRTV-----
fam1_22/1-553 -----
fam1_23/1-651 FRTVHEQDITYDFVAIQASDFRCQCKHDVPIFDECRSKYAI GDTVLVREWN-----
fam1_24/1-536 -----
fam1_25/1-349 -----

```

```

MgDNMT_deRIP/160-1440 AQIIIEFDTQKRRVVL RQFLRKSEQTGAKARPNELVLT DVVFDKACDKIIRKCHVRCFRQ
S1_Dim2/1-1281 AQIIIEFDTQKRRVVL RQFLRKSEQTGAKARPNELVLT DVVFDKACDKIIRKCHVRCFRQ
fam1_01/1-1281 AXIIIEFDTXKRRVVL RXFLRKSEQTGAKARPNELVLT DVVFNKAYDKIIRKCYMRCFRX

```

```

fam1_02/1-1281      AQIIIEFDTXKRRVVLRXFLRKSEXQIGAKARPNEVLIDVVFVNKACNKILQKCYIRYFRX
fam1_03/1-1282      AXIIKFDTXKRRVVLRLQFLRKSEQXTGAKARPNEVLTDVVFDKACDKIIRKCYVRYFRX
fam1_04/1-1282      AXIIKFDTXKRRVVLRLXFLRKSEQXTGAKARLNEVLTDVVFDKACDKIIRKCHMRYFRX
fam1_05/1-582      -----
fam1_06/1-156      AQIIIEFDTQKRRVVLRLQFLRKSEQQTGAKARPNEVLTDVVFDKACDKIIRKCYVRCFRQ
fam1_08/1-1281      AQIIIEFDTXKRRVVLRLXFLRKSEQQTGAKARPNEVLTDVVFNKAYDKIIRKCHIYCFRK
fam1_09/1-1281      AQIIIEFDTXKRRVVLRLXFLRKSEQQTGAKARPNEVLTDVVFNKAYDKIIRKCHIYCFRK
fam1_10/1-1281      TQIIIEFDTQKRRVVLRLQFLRKSEQQIGAKARPNEVLTDVVFNKACDKIIRKCYMRCFRQ
fam1_11/1-1281      AQIIIEFDTQKRRVVLRLQFLRKSEQXIGAKARPNEVLTDVVFDKACDKIIRKCYVRCFRX
fam1_12/1-1281      TQIIIEFDTQKRRVVLRLQFLRKSEQXTGAKARPNEVLTDVVFDKACDKIIRKCYVRCFRQ
fam1_13/1-697      TQIIIEFDTQKRRVVLRLQFLRKSEQQTGAKARPNEVL-----
fam1_14/1-1281      TQIIIEFDTQKRRVVLRLXFLQKSEQQTGAKARPNEVLTDVVFNKACDKIIRKCYMRCFRQ
fam1_15/1-1256      TQIIIEFDTXKRRVVLRLQFLQKSEQQIGAKARPNEVLTDVVFDKACDKIIRKCYVRCFRQ
fam1_16/1-1281      AQIIIEFDTQKRRVVLRLQFLQKSEQQMGAKARPNEVLTDVVFDKVDKIIRKCYVRCFRQ
fam1_17/1-651      -----
fam1_18/1-1281      TXIIIEFDTQKRRVVLRLQFLRKSEQQTGAKARPNEVLTDVVFNKACDKIIRKCYVRCFRQ
fam1_19_20/1-1281  AXIIIEFDTXKRRVVLRLXFLXSKQXIGAKARPNEVLTDVVFNKACDKIIRKCYMRCFRX
fam1_21/1-607      -----
fam1_22/1-553      -----
fam1_23/1-651      -----
fam1_24/1-536      -----
fam1_25/1-349      -----

MgDNMT_deRIP/160-1440 DQIDSGDIVTPYDRDGTGDFFYLAEDLSAARDRFPPMKEGWNPNMHYGEKMGMSLFYGG
S1_Dim2/1-1281      DQIDSGDIVTPYDRDGTGDFFYLAEDLSAARDGFPPMKEGWNPNVHYGEKMGMSLFCGG
fam1_01/1-1281      DQIDSRDIVLYDRDRTSNFFYLAKDLSVXDSFLPIKEGXNPNMHYGEKIVGISLFYSS
fam1_02/1-1281      DXIDSGDIVTLYDRDGTSDFFYLAKDLSVVRDSFLLIKEGXNPNMHYSEKIVGISLFCGS
fam1_03/1-1282      DQIDSRDIITLYDRDGTGDFFYLAEDLLVARDSFPPMKEGXNPNMYSEKIVGISLFYGS
fam1_04/1-1282      DQIDSRDIITLYDRDGTGDFFYLAEDLLVARDSFPPMKEGXNPNMYSEKIVGISLFYGS
fam1_05/1-582      -----
fam1_06/1-156      DQIDSGDIVTPYDR-----
fam1_08/1-1281      DXIDSRDIVTLYDRDGTSDFFYLAEADLSAARDSFLLIKEGWNPNIHSEKMGMSLFYGS
fam1_09/1-1281      DXIDSRDIVTLYDRDGTSDFFYLAEADLSAARDSFLLIKEGWNPNIHSEKMGMSLFYGS
fam1_10/1-1281      DQIDSRDIVTLYDXDGTSDFFYLAEALSAAQDSFLLMKEGWNPNMHYGEKMGMSLFYGS
fam1_11/1-1281      DQIDSGDIVTPYDQDGTGDFFYLAEDLSVVRDSFLLMKEGXNPNMHYGEKIVGISLFYGS
fam1_12/1-1281      DQIDSGDIVTLYDRDGTGDFFYLAELSAAQDGFPPMKEGWNPNMHGKIVGMSLFCGS
fam1_13/1-697      -----
fam1_14/1-1281      DQIDSGDIVTLYDRDGTGDFFYLAELSAVQDSFLPIKEGXNPNMHYGEKIVRISLFYGS
fam1_15/1-1256      DXIDSGDIVTLYDRDGTGDFFYLAEDLSTAQDSFLLIKEGXNPNMHYSEKMGMSLFYGS
fam1_16/1-1281      DXIDSGDIVTPYDRDGTGDFFYLAEDLSAARDSFPPMKEGWNPNMHGKIVGISLFCGS
fam1_17/1-651      -----
fam1_18/1-1281      DQIDSGDIVTLYNRDRTGDFFYLAEDLSVARDSFPLIKEGWNPNMHYGEKMGMSLFYGG
fam1_19_20/1-1281  DXIDSRDIVTLYDRDGTSDFFYLAEADLLVVRDRFLPIKEGXNPNMYSEKIVGISLFYGS
fam1_21/1-607      -----
fam1_22/1-553      -----
fam1_23/1-651      -----
fam1_24/1-536      -----DLVVRDSFLPIKEGWNPNMYSEKIVGISLFYGG
fam1_25/1-349      -----

MgDNMT_deRIP/160-1440 GTFDRLGLEEAGIARFKYAADWNEHALHSHRANVRRPDMEYFLGLVNDLLARAMAGDPKP
S1_Dim2/1-1281      GTFDRLGLEEAGIARFKYAVDWNEHALHSHRANVRRPDQMEYFLGSVNDLLARAMVGDSKS
fam1_01/1-1281      RTFNYSLEEAGIARFKYTADWNKYALYSYRANVRRPDRIEYFLGLVNDLLARAIAGDLKL
fam1_02/1-1281      RTFNCSLEEAGIARFKYAADWNKHALYSYRANVRRPNRIEYFLGLVNDLLTRAIAGDLIS
fam1_03/1-1282      RTFNRSLEEAGIARFKYTADXNKYTLHSYXANVRRPDRMEYFLGLVNDLLARAIAGDPKS
fam1_04/1-1282      RTFNRSLEEAGIARFKYTADXNKHTLHSYXANVRRPDRMEYFLGLVNDLLARAIAGDPKS
fam1_05/1-582      -----
fam1_06/1-156      -----
fam1_08/1-1281      RTFNCSLEEAGIARFKYTADXNKHTLHSYQANVRRPDRIEYFLGSVNDLLTRAIAGDLMS
fam1_09/1-1281      RTFNCSLEEAGIARFKYTADXNKHTLHSYQANVRRPDRIEYFLGSVNDLLTRAIAGDLMS
fam1_10/1-1281      RTFNRSLEEAGIARFKYAADXNKHTLYSYRANVRRPDRIEYFLGLVNDLLARAIAGDLKS
fam1_11/1-1281      RTFNRSLEEAGIARFKYAADWNEHALYSYRANVRRPDRMEYFLGSVNDLLTRAMAGDLIS
fam1_12/1-1281      GTFDRSLEEAGIARFKYAADWNEHALHSHRANVRRPDRMEYFLGSVNDLLARAMAGDPKS
fam1_13/1-697      -----
fam1_14/1-1281      RTFNCSLEEAGIARFKYAADWNKHALHSHRANVRRPDRMEYFLGLVNDLLPRAIAGDPKS
fam1_15/1-1256      RTFNRSLEEAGIARFKYAADWNKHTLHSYXANVRRPDRIEYFLGLVNDLLARAMAGDLKS
fam1_16/1-1281      RTFDHSLEEAGIARFKYAADXNKHTLHSYRANVRRPDRIEYFLGSVNDLLARAMAGDLMS
fam1_17/1-651      -----
fam1_18/1-1281      GTFDCGLEEAGIARFKYAADXNEHALHSYRANVRRPDRMEYFLGLVNDLLARAIAGDPKS
fam1_19_20/1-1281  RTFNRSLEEASIAKLKYAVDXNEHALHSYANVRRPDRIEYFLGLVNDLLARAIAGDLKL
fam1_21/1-607      -----
fam1_22/1-553      -----
fam1_23/1-651      -----
fam1_24/1-536      RTFNYSLEEAGIARFKYAADWNKHALHSYXANVRRPDRIKYFLSLVNDLLTRAIAGDLKS

```

```

fam1_25/1-349 -----
MgDNMT_deRIP/160-1440 NIARPGDIQILAGGSPCPGFSMLQLFKLSEQSKQNASLVASIVSFVDFYVPRYFLLKNVV
S1_Dim2/1-1281 NIAQPGDIQILAGGSPCPGFSMLQLFKLSEQSKQNASLVASIVSFVDFYVPRYFLLKNVV
fam1_01/1-1281 NIARPGDIYILIGGSLCPRFSMLQLFKLSEQSKQNASLVASIVSFVDFYVPRYFLLKNVV
fam1_02/1-1281 NIAXLGDIIYILTGGSLCPGFSILQLFKLSEQSKQNASLVASIMLFVDFYVPRYFLLKNVV
fam1_03/1-1282 NIAXPGDIYILIGGSPCPGFSMLQLFKLSEQSKXNASLVASIVLFDYVPRYFLLKNVI
fam1_04/1-1282 NIAXPGDIYILIGGSPCPGFSMLQLFKLSEQSKXNASLVASIIILFDYVPRYFLLKNVI
fam1_05/1-582 -----
fam1_06/1-156 -----
fam1_08/1-1281 NIAQPGDIYILTGGSPYPGFSMLQLFKLSEQSKQNASLVASIVSFVDFYVPRYFLLKNVV
fam1_09/1-1281 NIAQPGDIYILTGGSPYPGFSMLQLFKLSEQSKQNASLVASIVSFVDFYVPRYFLLKNVV
fam1_10/1-1281 NIARPGDIYILTGGSPCPGFSMLQLFKLSEQSKQNTSLVASIVLFDYVPRYFLLKNVV
fam1_11/1-1281 NIARPGDIHILTGGSPYPGFSILXLFKLSEQSKXNASLVASIVSFVDFYVPRYFLLKNVV
fam1_12/1-1281 NIARPGDIHILTGGSPCPGFSMLQLFKLSEQSKXNASLVASIVLFDYVPRYFLLKNVV
fam1_13/1-697 -----
fam1_14/1-1281 NIARPGDIHILTGGSPYPGFSMLQLFKLSEQSKQNASLVASIVSFVDFYVPRYFLLKNVI
fam1_15/1-1256 NIAQPGDIHILTGGSPCPGFSMLQLFKLSEQSKQNASLVASIVLFDYVPRYFLLKNVV
fam1_16/1-1281 NIARPGDIHILTGGSPCPGFSMLQLFKLSEQSKQNALLVASIVSFVDFYVPRYFLLKNVV
fam1_17/1-651 -----
fam1_18/1-1281 NIARPGDIHILTGGSPCPGFSMLQLFKLSEQSKQNASLVASIVSFVDFYVPRYFLLKNVV
fam1_19_20/1-1281 NIAXPGDIXILVGGSLCPRFLILQLFKLSKQSKQNASLVASIVSFVDFYVPRYFLLKNVV
fam1_21/1-607 -----
fam1_22/1-553 -----
fam1_23/1-651 -----
fam1_24/1-536 NIARLGDIIHILTGGSLCPGFSMLXLFKLSKQSKQNASLVASIVSFVDFYVPRYFLLKNVV
fam1_25/1-349 -----

MgDNMT_deRIP/160-1440 TITAGMGPKNKDQNVFS-QIVAALVALGYQVQQFLGDAWSHGSPQQRSRVFIAASAPGLPL
S1_Dim2/1-1281 TMTAGMGPKNKDQNVFS-QIVAALVALGYQVQQFLGDAWSYGSPQQRSRVFIVASAPGLPP
fam1_01/1-1281 TITAGIGPNKDQNVFL-QIVTALVALSYXVXXFLGDAWSYGLXQRLXVFISTASAPSLPL
fam1_02/1-1281 TITAGIGPNKDXNVFL-XIVTALVALSYXVXQFLGDAXSYGSLQQLRVFISASAPGLPL
fam1_03/1-1282 TITAGMGPKNDXNVFLLQIVTALVALSYXVXQFLGDAWSHSLXQRLRVFISASAPSLLL
fam1_04/1-1282 TITAGIGPNKDXNVFSLQIVTALVALSYXVXQFLGDAWSHGLXQRLRVFISASAPSLLL
fam1_05/1-582 -----
fam1_06/1-156 -----
fam1_08/1-1281 TITAGIGPNKDQNVFS-QIVAALVALSYXVXQFLGDAWSHSSPQXQSRVFISTASAPSLPL
fam1_09/1-1281 TITAGIGPNKDQNVFS-QIVAALVALSYXVXQFLGDAWSHSSPQXQSRVFISTASAPSLPL
fam1_10/1-1281 TITAGIGPNKDQNVFL-QIVAALVALSYXVXQFLGDAWSHGSPQQRSRVFISASAPGLLL
fam1_11/1-1281 IITAGIGPNKDQNVFS-QIVAALVALSYXVQFLGDAWSHGSPXQSRVFISTASAPGLLP
fam1_12/1-1281 TMTAGMGPKNKDQNVFL-XIVAALVALGYQVQQFLGDAWSHGSPQQRSQVFISASAPGLPL
fam1_13/1-697 -----
fam1_14/1-1281 TITAGIGPNKDQNVFL-QIVAALVALSYQVQQFLGDAXSYGSLQQLRVFISTASAPSLPL
fam1_15/1-1256 TITAGIGPNKDQNVFS-QIVAALVALGYQVXQFLGDAWSYSSLXQRLRVFISASAPSLPL
fam1_16/1-1281 TITAEIGPNKDQNVFS-QIVAALVALGYXVQFLGDAXSHGSPQQLRVFISASAPGLPL
fam1_17/1-651 -----
fam1_18/1-1281 TITAGMGPKNKDQNVFS-QIVAALVALSYXVXQFLGDAWSHGSPQQRXSVFISASAPGLLL
fam1_19_20/1-1281 TITAGIGPNKDXNMFs-XIVAALVALGYXVXQFLGDAXSYGLXXRLRVFIVASAPGLLL
fam1_21/1-607 -----
fam1_22/1-553 -----
fam1_23/1-651 -----
fam1_24/1-536 TITAGIGLNKDQNVFS-QIVTALVALSYXVQFLGDAXSHGSPQXRLRVFISTASAPSLLL
fam1_25/1-349 -----

MgDNMT_deRIP/160-1440 LSPIAHSHSYPNMSFRSRALGLSSNGLPFGSRRDEFTAFHVTPAQACADLPDIGAQPQ
S1_Dim2/1-1281 LSPLAHSHSHPNMSFRTRALGLSSNGLPFGSRRDEFTAFPHVTPAQACADLPDIGAQPQ
fam1_01/1-1281 LSLIPYSYSYLNISFRSRALSLSSNGLPFGSRKDEFTTFLYITPAXAYADLLDIGDAXLX
fam1_02/1-1281 LSLIPYSYSYLNISFRSRALSLSSNGLPFGSCKDEFTTFLYVTLAXACADLLDIGDAXLX
fam1_03/1-1282 LSLIPYSYSYLNISFRSCALGLSSNGLPFGSRKDEFTTFLYITPAXAYADLLDIRDAQPO
fam1_04/1-1282 LLLIPYSYSYLNISFRSCALGLSSNGLPFGSRKDEFTTFLYVTPAXAYADLLDIRDAQPO
fam1_05/1-582 -----
fam1_06/1-156 -----
fam1_08/1-1281 LSLIPYSYSYLNISFRSRVLSFSSNGLPFGSRKDEFTTFLYVTPAXAYADLLDIGDAQPX
fam1_09/1-1281 LLLIPYSYSYPNISFRSRALSFSSNGLPFGSRKDEFTTFLHVTPAXAYADLLDIGDAQPO
fam1_10/1-1281 LSLIPYSYSYLNISFRSRALGLSSNGLPFGSRKDEFTTFLYVTPAQACADLLDIGDAQPO
fam1_11/1-1281 LSLIPYSHSHPNMSFRSRTLSSNGLPFGSRKDEFTTFLYVTPAXAYADLLDIGDTQPO
fam1_12/1-1281 LSLIPHSHSHPNMSFRSRALSLSSNGLPFGSCKDEFTAFLYVTPAQACADLPDIGAQPQ
fam1_13/1-697 -----
fam1_14/1-1281 LSLIPYSHSYPNISFRSRALGLSSNGLPFGSRKDEFTTFLHVTPAXACADLPDIGAQLQ
fam1_15/1-1256 LSLIPYSYSYLNMSFRSRALGLSSNGLPFGSRKDEFTTFLYVTPAXACADLPDIGDTQPO
fam1_16/1-1281 LSLIPHYSYSYPNISFRSRALSLSSNGLPFGSRKDEFTTFLYVTPAQACADLPDIGAQLQ
fam1_17/1-651 -----
fam1_18/1-1281 LLPIPHYSYSYLNISFRSRALGLSSNGLPFGSRKDEFTAFLYVTPAQACADLLDIGDAQPO

```

```
fam1_19_20/1-1281  LSLIAYSYSYLNISFRSRALGLSSNGLPFGSRRDEFTTFLYVTPAXAYADLLDIGDAPX
fam1_21/1-607 -----
fam1_22/1-553 -----
fam1_23/1-651 -----
fam1_24/1-536  LSLIPYSYSYLNISFRSRALSLSSNGLPFGSRKDEFTTFLYVTPAXACADLLDIGDAPX
fam1_25/1-349 -----
```

```
MgDNMT_deRIP/160-1440 LCPQFPDYRTMSLESATNRTRMAAVPTRPHGVGLMQAKLLGLLRGEPLTFCDNVNATRRQ
S1_Dim2/1-1281 LCPQFPDHRMSLESATNRTRMAAVPTRPHGLGLMQAKLLGLLRGEPLTFCDNVNATRRQ
fam1_01/1-1281 LCPQFLDYXTILLELVNTRTRIAAMPTRPYRIGLIXAKLLGLLRGEPLTFYDNVNATRRX
fam1_02/1-1281 LCLXFLDYRTMSLELATNRTRIVAVPTRPYRIGLIXAKLLGLLCGEPLTFCDNINAIRRX
fam1_03/1-1282 LCLXFPDYRTMSLESATNRIRMAAVPTRLYRIGLIXAKLLGLLRGEPLTFCDNVNATRRQ
fam1_04/1-1282 LCLXFPDYRTMSLESATNXIQIAAVPTRPYRIGLIXAKLLGLLRGEPLTFCDNVNATRRQ
fam1_05/1-582 -----
fam1_06/1-156 -----
fam1_08/1-1281 LCLQFPDHRMSLESATNXIQIAAMPTRPYGMGLMXAKLLGLLRGEPLTFCDNINATRRQ
fam1_09/1-1281 LCLQFPDHRMSLESATNXIQIAAMPTRPHRMGLIQAKLLGLLRGEPLTFCDNINATRRQ
fam1_10/1-1281 LCLXFLDYRTMSLESATNRTRIAAVPTRPYRIGLIXAKLLGLLRGEPLTFYDNVNATRRQ
fam1_11/1-1281 LCLXFPDHRMSLESATNRTRIAAMPTRLYRIGLIXAKLLGLLRGEPLTFCDNVNATCRX
fam1_12/1-1281 LCPQFPDHRMSLESATNRTRMAAVPTRPYGMGLIQAKLLGLLRGEPLTFCDNVNATRRQ
fam1_13/1-697 -----
fam1_14/1-1281 LCLQFPDYRTMSLESATNQTRMAAVPTRPYRIGLIXAKLLGLLRGEPLTFCDNVNATRRX
fam1_15/1-1256 LCPQFLDYRTMLLESATNRTRIAAVPTRPHRIGLIXAKLLGLLRGEPLTFCDNVNATRRQ
fam1_16/1-1281 LCPQFPDYRTMLLESATNRIRMAAIPTRPHRIGLIXAKLLGLLRGEPLTFCDNVNATRRQ
fam1_17/1-651 -----
fam1_18/1-1281 LCPQFLDYRTMSLESATNRTRMAAVPTRPYRIGLIXAKLLGLLRGEPLTFCDNVNATCRQ
fam1_19_20/1-1281 LCPQFLDYRTMSLESATNRIRIAAIPTRLYRVGLIXAKLLGLLRGEPLTFCDNINATCRQ
fam1_21/1-607 -----
fam1_22/1-553 -----
fam1_23/1-651 -----
fam1_24/1-536 LCLXFLDYRTMSLESATNXIRIAAMPTRPHRIGLIXAKLLGLLCGEPLTFCDNVNATRRQ
fam1_25/1-349 -----
```

```
MgDNMT_deRIP/160-1440 ENSRMFSRMFPNGLFPTIVTKTAIGDSINRRFMHWSQNRIWTVIEARRAGYXDWEVLIG
S1_Dim2/1-1281 ENSKMYSRIFPNGLFPTIVTKTAIGDINGRFMHWSQNRIWTVMEARRAQGYQDWEVLIG
fam1_01/1-1281 ENSRMFSRMFLNSLFLTIVTKTAIRDSINRRFMHXSXNRIXTVIEARRAGYXDWEVLIG
fam1_02/1-1281 ENSRMFSRMFLNSLFLTIVTKTAIRDSINRRFMHXSXNRIXTVIEARRAGYXDWEVLIG
fam1_03/1-1282 ENSRMFSRMFLNSLFLTIVTKTAIGDSINRRFIYXSXNRIXTVIEARRAGYXDWEVLIG
fam1_04/1-1282 ENSRMFSRMFLNSLFLTIVTKTAIGDSINRRFIYXSXNRIXTVIEARRAGYXDWEVLIG
fam1_05/1-582 -----
fam1_06/1-156 -----
fam1_08/1-1281 ENSRMFSRMFPNSLFLTIVTKTAIGDSINRRFIHXSQNRIWTVIEARRAQGYXDWEVLIG
fam1_09/1-1281 ENSRMFSQMFPNSLFLTIVTKTAIGDSINRRFIHXSQNRIWTVIEARRAQGYXDWEVLIG
fam1_10/1-1281 ENSRMFSRMFLNGLFPTIVTKTAIGDSINRRFMHXSXNRIWTVIEARRAGYXDWEVLIG
fam1_11/1-1281 ENSRMFSRMFLNSLFLTIVTKTAIGDSINRRFMHXSXNRIWTVIEARRAQGYXDWEVLIG
fam1_12/1-1281 ENSRMFSRMFPNGLFPTIVTKTAIGDSINRRFMHWSQNRIWTVIEARRAQGYXDWEVLIG
fam1_13/1-697 -----
fam1_14/1-1281 ENSRMFSRMFLNSLFLTIVTKTAIGDSINRRFMHXSXNRIWTVIEARRAGYXDWEVLIG
fam1_15/1-1256 ENSRMFSRMFPNSLFLTIVTKTAIGDSINRRFMHXSXNRIWTVIEARRAGYXDWEVLIG
fam1_16/1-1281 ENSRMFSRMFLNSLFLTIVTKTAIGDSINRRFMHXSXNRIWTVMEARRAGYXDWEVLIG
fam1_17/1-651 -----
fam1_18/1-1281 ENSRMFSRMFPNGLFPTIVTKTAIGDSINRCFMHWSQNRIWTVIEARRAGYXDWEVLIG
fam1_19_20/1-1281 ENSRMFLRMFLNSLFLTIVTKTAIGDSINRRFIYXSXNRIXTVIEARRAGYXDWEVLIG
fam1_21/1-607 -----
fam1_22/1-553 -----
fam1_23/1-651 -----
fam1_24/1-536 ENSRMFSRMFLNSLFLTIVTKTAIGDSINRRFMHXSXNRIXTVIEARRAGYXDWEVLIG
fam1_25/1-349 -----
```

```
MgDNMT_deRIP/160-1440 SPAXQHKIVGNSVNRKVALVLGLSLRESXERARSQLLFPRASEIRDEHVVEEIGTDITAE
S1_Dim2/1-1281 SPAQQHKIVGNSVDRKVALVLGLSLRESWEQARSQSLFPRASEIRDEHVVEEEMDTKITAE
fam1_01/1-1281 SPAXQHKIVGNSINRKVALVLGLSLRESXEXARSXLLFLRASKIRNEHVVEEIGTDITIE
fam1_02/1-1281 SPAXQYKIVGNSVNRKVALVLGLSLRESXEXARSXLLFLRASKIRNEHVVEEIGTDITIE
fam1_03/1-1282 SPTXXHKIVGNSVNRKVALVLGLSLRESXEQARSXLLFLRASEIRNEHVVEEIGTDITIE
fam1_04/1-1282 SPTXXHKIVGNSVNRKVALVLGLSLRESXEQARSXLLFLRASEIRNEHVVEEIGTDITIE
fam1_05/1-582 -----
fam1_06/1-156 -----
fam1_08/1-1281 SPTXQHKIVGNSVNRKVALVLGLSLRESXEQARSQLLFPRAASKIRNEHVEKEIGTDITVE
fam1_09/1-1281 SPTXQHKIVGNSVNRKVALVLGLSLRESXEQARSQLLFPRAASKIRNEHVEKEIGTDITVE
fam1_10/1-1281 SPAXXHKIVGNSVNRKVALMLGLSLRELXEXARSXLLFLRASKIRNEHVVEEIGTDITAE
fam1_11/1-1281 SPAXQHKIVGNSINRKVALVLGLSLRESXEQARSQLLFLRASKIRNEHVVEEIGTDITVE
fam1_12/1-1281 SPAQQHKIVGNSVDRKVALVLGLSLRESXEQARSQLLFPRASEIRDEHVVEEEMGTNITVE
fam1_13/1-697 -----
```

```

fam1_14/1-1281      SPAXQHKIVGNSVNRKVALVLGLSLRESXEQARSQLLFLRASKIRNEHVVEEIGTDITTE
fam1_15/1-1256      SPAXQHKIVGNSVNRKVALVLGLSLRESXEQARSQLLFPRASKIRNEHVVEEIGTDITIE
fam1_16/1-1281      SPAXQHKIVGNSVNRKVALVLGLSLRESXEQARLQLLFLRVSKIRDEHVVEEIGTDITAE
fam1_17/1-651      -----
fam1_18/1-1281      SPAXQHKIVGNSVNRKVALVLGLSLRESXEXARLQLLFPVSEIRDEHVVEEIGTDITVE
fam1_19_20/1-1281   SPAXXYKIVGNSVNRKVALVLGLSLRESXEQARSXLLFLRASEIRNEHIKEEIGTDITIE
fam1_21/1-607      -----
fam1_22/1-553      -----
fam1_23/1-651      -----
fam1_24/1-536      SPAXXHKIVGNSVNRKVALVLGLSLRESXEXARLXLLFLRASKICNEYVKEEIGTDITIE
fam1_25/1-349      -----

MgDNMT_deRIP/160-1440 ASQPTTPTKDEDVEIEGDSSSVATRGALKEEEEQSRALMAPSTPQVGPSARLNIEELDNRT
S1_Dim2/1-1281      VARPATTTDDEDVEIEGFATSVATGGALGEAEQSRAPMAPSTPQVGPSARLNIEELDNRT
fam1_01/1-1281      ASQPIPTPKDKDVEIEGNSSSVATRGTLKEEEQLRALIVPSTPXVSPSARLNIEELDNRT
fam1_02/1-1281      ASQPTTPTKDEDVEIEGNLSSSVATRGTLKEEEQLRALIVLSTPXVSPSARLNIEELDNRT
fam1_03/1-1282      ASQPATPTKDEDVEIEGDSSSVATRGTLKEEEQLRALIAPSTXVSPSARLNIEELDNRT
fam1_04/1-1282      ASQPTTPTKDEDVEIEGDSSSVATRGTLKEEEQLRALIAPSTPXVSPSARLNIEELDNRT
fam1_05/1-582      -----
fam1_06/1-156      -----
fam1_08/1-1281      ASQPTTPTKDEDVEIEGNSSSVATRGALKEEEQLRALIALSTPXVSPSARLNIEELDNRT
fam1_09/1-1281      ASQPTTPTKDEDVEIEGNSSSVATRGALKEEEQLRALIALSTPXVSPSARLNIEELDNRT
fam1_10/1-1281      ASQPIPTPKDEDVEIEGNSSSVATRGALKEEEQLRALIALSTPXVSPSARLNIEELDNRT
fam1_11/1-1281      ALQPTTPTKDEDVEIEGNSSSVATRGALKEEEQLRALIAPSTPXVSPSARLNIEELDNRT
fam1_12/1-1281      ASXPATPTKDEDVEIEGNSSSVATRGALKEEEQSRAPMAPSTPXVSPSARLNIEELDNRT
fam1_13/1-697      -----
fam1_14/1-1281      ALXPTTPTKDEDVEIEGNSSSVATRGALKEEEQSRALIALSTPXVSPSARLNIEELDNRT
fam1_15/1-1256      ASQPATPTKDEDVEIEGNSSSVATRGALKEEEQSRALIVLTPQVGPSARLNIEELDNRT
fam1_16/1-1281      ASQPTTPTKDEDVEIEGNSSSVATRGALKEEEQSRALIALSTPXVSPSARLNIEELDNRT
fam1_17/1-651      -----
fam1_18/1-1281      ASXPATPTKDEDVEIEGNSSSVATRGALKEEEQSRALMAPSTPQVGPSARLNIEELDNRT
fam1_19_20/1-1281   ASXPITPTKDEDVEIEGNLSSSVATRGALKEEEQLRALIALSIPXVSPSARLNIEELDNRT
fam1_21/1-607      -----
fam1_22/1-553      -----
fam1_23/1-651      -----
fam1_24/1-536      ASXPITPTKDEDVEIEGNLSSSVATRGTLKEEEQLRALIVLSTPXVSPSARLNIEELDNRT
fam1_25/1-349      -----

MgDNMT_deRIP/160-1440 AKMLLFMPPEEVDVNSLEGGEF
S1_Dim2/1-1281      AKMLLFMPGEVGDVNSLEGGEF
fam1_01/1-1281      AKMLLFMLEEVDVNSLEGGEF
fam1_02/1-1281      AKMLLFMPPEEVDVNSLEGGEF
fam1_03/1-1282      AKMLLFMPPEEVDVNSLEGGEF
fam1_04/1-1282      AKMLLFMPPEEVDVNSLEGGEF
fam1_05/1-582      -----
fam1_06/1-156      -----
fam1_08/1-1281      AKMLLFMPPEEVDVNSLEGGEF
fam1_09/1-1281      AKMLLFMPPEEVDVNSLEGGEF
fam1_10/1-1281      AKMLLFMPPEEVDVNSLEGGEF
fam1_11/1-1281      AKMLLFMPPEEVDVNSLEGGEF
fam1_12/1-1281      AKMLLFMPPEEVDVNSLEGGEF
fam1_13/1-697      -----
fam1_14/1-1281      AKMLLFMLEEVDVNSLEGGEF
fam1_15/1-1256      -----
fam1_16/1-1281      AKMLLFMPPEEVDVNSLEGGEF
fam1_17/1-651      -----
fam1_18/1-1281      AKMLLFMPPEEVDVNSLEGGEF
fam1_19_20/1-1281   AKMLLFMPPEEVDVNSLEGGEF
fam1_21/1-607      -----
fam1_22/1-553      -----
fam1_23/1-651      -----
fam1_24/1-536      AKMLLFMPPEEVDVNSLEGGEF
fam1_25/1-349      -----

```

FIGURE S3.—Protein alignments of MgDNMT sequences in the genome of *Mycosphaerella graminicola*. Protein alignments of 23 MgDNMT sequences were obtained using ClustalX. The ‘deRIPed’ MgDNMT copy and species S1 DIM-2-like sequence was also included in the alignment. Stop codons in the protein sequences have been replaced by the character ‘X’.

FIGURE S4.—Alignment of the MgDNMT protein of *Mycosphaerella graminicola* to DIM-2-like proteins in other fungi. Protein alignment of ‘deRIPed’ MgDNMT sequence to DIM-2-like proteins from other fungal species. Intensity of the blue color increases with the percentage identity between the sequences. Species corresponding to the protein names in the alignment: *M. graminicola*, chromosome 6 ‘deRIPed’ MgDNMT; *M. fijiensis*, MfDim2-like; S1, S1-Dim2; CIMG_01762, *Coccidioides immitis*; PODANSg1750, *Podospora anserina*; FG10766, *Gibberella zeae*; BC1G_12419, *Botryotinia fuckeliana*; MGG_00889, *Magnaporthe oryzae*; SNOG_03039, *Phaeosphaeria nodorum*; PTRG_02280, *Pyrenophora tritici-repentis*; SS1G_07976, *Sclerotinia sclerotiorum*; NCU02247, *Neurospora crassa*.

TABLE S1**MgDNMT sequences and their co-ordinates in the *Mycosphaerella graminicola* genome**

Number	Family_Element	Chromosome	Start	End	Length
1	1_1	1	6049446	6053867	4421
2	1_2	1	6079895	6084316	4421
3	1_3	1	28287	32365	4078
4	1_4	1	63897	67977	4080
5	1_5	1	1409	3629	2220
6	1_6	1	152	604	452
7 ^a	1_7	1	46196	46543	347
8	1_8	2	3839548	3843966	4418
9	1_9	2	3816897	3821315	4418
10	1_10	15	618972	623393	4421
11	1_11	17	51582	56003	4421
12	1_12	17	25130	29546	4416
13	1_13	17	574208	576771	2563
14	1_14	5	2822213	2826631	4418
15	1_15	5	2856701	2860944	4243
16	1_16	21	27289	31710	4421
17	1_17	21	161	2589	2428
18	1_18	6	17912	22330	4418
19 ^b	1_19	6	1267761	1270603	2842
20 ^b	1_20	6	1266188	1267676	1488
21	1_21	12	1442424	1444694	2270
22	1_22	16	500	2635	2135
23	1_23	9	161	2589	2428
24	1_24	8	7701	9410	1709
25	1_25	18	565163	566640	1477
26 ^a	1_26	7	2636392	2636634	242

^aLength less than 400bp. Not included in the DNA or protein alignments.

^bNon-telomeric copy on chromosome 6, split in two.