MutS-Homolog2 silencing generates tetraploid meiocytes in tomato (*Solanum lycopersicum*)

Supriya Sarma¹, Arun Kumar Pandey¹, Maruthachalam Ravi², Yellamaraju Sreelakshmi¹, Rameshwar Sharma¹*

¹Repository of Tomato Genomics Resources, Department of Plant Sciences, School of Life Sciences, University of Hyderabad, Hyderabad-500046, India

²School of Biology, Indian Institute of Science Education and Research, Thiruvananthapuram, CET Campus, Thiruvananthapuram 695016, Kerala, India

***Corresponding author:** Repository of Tomato Genomics Resources, Department of Plant Sciences, School of Life Sciences, University of Hyderabad, Hyderabad-500046, India. <u>rameshwar.sharma@gmail.com</u>, Tel: +91-40-23010514, Fax: +91-40-23010120

Authorsemailaddresses:supu.megha@gmail.com(S.Sarma),pandeyarun80@gmail.com(A.K.Pandey),ravi@iisertvm.ac.in(M.Ravi),syellamaraju@gmail.com(Y.Sreelakshmi),rameshwar.sharma@gmail.com(R.Sharma)

Running title: *MSH2* silencing induces tetraploidy. Number of tables: Two Number of figures: Six Supplementary figures: Eight Supplementary tables: Four

Keywords: MutSHomolog2, meiosis, cytokinesis, mitosis, polyploidy, tetraploid meiocytes, tomato, *Solanum lycopersicum*.

SUMMARY

MSH2 is the core protein of MutS-homolog family involved in recognition and repair of the errors in the DNA. While other members of MutS-homolog family reportedly regulate mitochondrial stability, meiosis, and fertility, MSH2 is believed to participate mainly in mismatch repair. The search for polymorphism in MSH2 sequence in tomato accessions revealed both synonymous and nonsynonymous SNPs; however, SIFT algorithm predicted that none of the SNPs influenced MSH2 protein function. The silencing of MSH2 gene expression by RNAi led to phenotypic abnormalities in highlysilenced lines, particularly in the stamens with highly reduced pollen formation. MSH2 silencing exacerbated formation of UV-B induced thymine dimers and blocked lightinduced repair of the dimers. The MSH2 silencing also affected the progression of male meiosis to a varying degree with either halt of meiosis at zygotene stage or formation of diploid tetrads. The immunostaining of male meiocytes with centromere localized CENPC (Centromere protein C) protein antibody showed the presence of 48 univalent along with 24 bivalent chromosomes suggesting abnormal tetraploid meiosis. The mitotic cells of root tips of silenced lines showed diploid nuclei but lacked intervening cell plates leading to cells with syncytial nuclei. Thus we speculate that tetraploid pollen mother cells may have arisen due to the fusion of syncytial nuclei before the onset of meiosis. It is likely that in addition to Mismatch repair (MMR), MSH2 may have an additional role in regulating ploidy stability.

INTRODUCTION

The survival of the living organisms is dependent on the precise and error-free replication of DNA. In addition, normal metabolic activity and environmental factors such as radiation also cause damage to the DNA. The fidelity of the DNA is ensured by a cohort of mechanisms that detects and eliminates mistakes occurring during DNA replication. At least five major repair pathways are known to operate in higher plants: base excision repair, nucleotide excision repair, mismatch repair (MMR) and double-strand break repair comprising of homologous recombination repair and nonhomologous end joining (Takata *et al.*, 1998; Guarné *et al.*, 2004; Reardon and Sancar, 2005; Jacobs and Schär, 2012). Among these, MMR is a major pathway that corrects base–base and insertion–deletion mismatches generated during DNA replication or mutagenesis (Dowen *et al.*, 2010; Srivatsan *et al.*, 2014).

MMR is a highly conserved pathway that exists in all living organisms. The MMR process consists of three key steps: mismatch recognition, excision, and resynthesis (Ban et al., 1999; Antony and Hingorani, 2003; Guarné et al., 2004). The first step of the mismatch recognition in eukaryotes is carried out by the homologs of prokaryotic MutS proteins, namely MSH protein subunits. There are eight homologs of MutS in eukaryotes: MSH1 to MSH8, of which MSH7 is found only in plants (Culligan and Hays, 2000) and MSH8 in Euglenozoa (Sachadyn, 2010). The MSH proteins recognize mismatches as heterodimers, where MutSa (MSH2-MSH6) repairs base-base mismatches or 1-2 nucleotide insertion-deletion loops (Acharya et al., 1996; Genschel et al., 1998), while MutS_β (MSH2-MSH3) recognizes larger, up to 14 nucleotide insertion-deletion loops (Modrich, 1991; Marti et al., 2002). Plants form an additional heterodimeric complex known as MutSy (MSH2–MSH7) (Culligan and Hays, 2000) which recognizes some base-base mismatches and reportedly plays a role in meiotic recombination (Lloyd *et al.*, 2007). The binding of MutS α/β to mismatched DNA initiates conformational changes in MutS recruiting MutL complex followed by activation of exonuclease1/PCNA. The gap generated by DNA excision is filled by action of DNA polymerase and DNA ligase.

It is reported that the other members of MSH2 family function beyond MMR. MSH1 is required for mitochondrial stability (Reenan and Kolodner, 1992) and disruption of MSH1 can influence male sterility in several species (Sandhu *et al.*, 2007;

Zhao *et al.*, 2016). The MSH4 and MSH5 form heterodimer between them without involving MSH2 and exclusively function in meiosis (Hollingsworth *et al.*, 1995; Schofield and Hsieh, 2003, Lu *et al.*, 2008; Wang *et al.*, 2016). The expression of MSH7 is required for wild-type level of fertility in barley (Lloyd *et al.*, 2007). MSH7 also plays a role in UV-B-induced DNA damage recognition and recombination in Arabidopsis (Lario *et al.*, 2015).

Among the MMR proteins, MSH2 is a common protein in three of the heterodimers formed in the plants. In conformity of the role of MSH2 protein in repair of replication errors, an Arabidopsis mutant defective in MSH2 shows microsatellite instability (Depeiges *et al.*, 2005). The *msh2* mutants in Moss show strong developmental defects, are sterile and have a mutator phenotype (Trouiller *et al.*, 2006). A similar phenotype was also observed in *Arabidopsis thaliana msh2* mutant lines albeit only after five generations (Hoffman *et al.*, 2004). Considering that suppression of MMR leads to an increase in natural mutation frequency, *MSH2* gene was silenced in *Nicotiana tabacum* and *N. plumbaginifolia* using an artificial microRNA (amiRNA). Though the transformed lines showed developmental and morphological abnormality, the mutant phenotype was not transmitted to subsequent generations (van Marcke and Angenon, 2013).

Among the interacting partners for MSH2; mutants are reported only for *MSH6* and *MSH7* in Arabidopsis. The *msh6* mutant does not display a visible phenotype (Lario *et al.*, 2011) however, *msh7* mutant shows an increase in meiotic recombination frequency (Lario *et al.*, 2015). The role of *MSH7* in meiotic recombination is also reported in tomato, where RNAi silencing of *MSH7* improves homeologous meiotic recombination with a chromosome from *S. lycopersicoides* (Tam *et al.*, 2011). Most likely the high increase in meiotic recombination is mediated by the MSH2-MSH7 complex as Arabidopsis *msh2* mutant also shows 40% increase in the meiotic crossover rates (Emmanuel *et al.*, 2006).

In this study, our aim was to isolate MSH2-defective mutants to obtain a hypermutable line to increase the efficiency of EMS-mutagenesis in tomato. We screened natural accessions of tomato and EMS-mutagenized lines for SNPs/mutation in tomato *MSH2* by EcoTILLING and TILLING respectively. Considering that little polymorphism was observed in *MSH2* in tomato accessions, we silenced *MSH2* gene

using *MSH2*-RNAi construct. We report that silencing of *MSH2* gene led to a developmental defect in tomato lines with abnormal flowers bearing tetraploid meiocytes in the anthers. The examination of root cells showed the formation of syncytial nuclei and defects in the cell plate formation. Our study indicates that *MSH2* may also have a role in the regulation of meiotic and mitotic cell divisions.

RESULTS

EcoTILLING of MSH2 gene showed little polymorphism in tomato accessions

Given the pivotal role of *MSH2* in MMR, we screened a total of 391 tomato accessions for polymorphism in the gene. A 2.7 Kb region (4-9 exon) of *MSH2* predicted to be deleterious by CODDLE software was chosen for EcoTILLING. Consistent with the essential nature of the gene, only 7 accessions showed polymorphism with a total of 13 SNPs with a frequency of 1.23 SNP/100 Kb. (Figure S1). Among the identified SNPs, 6 were non-synonymous, and 7 were synonymous (Table 1). The likelihood of non-synonymous SNPs affecting protein function was examined by SIFT software. Since none of the SNP showed a SIFT score <0.05, these SNPs were predicted to be non-deleterious in nature. A parallel screening of 1243 bp region of *MSH2* gene covering 4th exon of 11304 tomato EMS-mutagenized lines by TILLING did not yield any mutation.

Generation of MSH2-RNAi lines

Considering recalcitrance to the mutagenesis and low polymorphism in *MSH2* gene, we silenced the gene by RNAi. A 301 bp region of *MSH2* gene with no significant homology to any other tomato gene was amplified and cloned into pHELLSGATE 12 vector (Figure S2). A total of ten independent transgenic lines bearing *MSH2*-RNAi construct were raised using *Agrobacterium*-mediated transformation. Southern blot analysis of T_0 lines revealed the presence of the transgene in five lines (Figure S3). In T_2 generation out of three lines, the lines $1T_21-11$ and $2T_27-5$ with strong MSH2 silencing showed a single transgene copy (Figure S4).

MMR transcript analysis in MSH2-RNAi lines using real-time PCR

To ascertain the magnitude of gene silencing, the level of *MSH2* transcript in transgenic lines was quantified by real-time PCR. Out of five Southern positive lines, the line $1T_21-11$ showed ~80% reduction of *MSH2* transcript level followed by line $2T_25-5$ and $2T_27-5$ compared to wild-type (WT) control (Figure 1). Line $3T_21-4$ showed moderate, and $5T_23-3$ showed no reduction in *MSH2* transcript levels. The reduction in the transcript level in line1T21-11 was not restricted to *MSH2* alone; it also

reduced the transcript levels of *MSH5* and *MSH7* albeit to a moderate but a significant extent. Since, *MSH2*, *MSH5*, and *MSH7* genes bear no significant sequence homology, the reduction in *MSH5* and *MSH7* transcript is most probably due to some indirect influence of *MSH2* transcript reduction.

MSH2 silencing exacerbates UV-B-mediated thymine dimer formation

In mammalian cells, the contribution of the MMR pathway to the UVB-induced DNA damage is well established (Narine *et al.*, 2007; Seifert *et al.*, 2008). Similarly, Arabidopsis *MSH2* T-DNA insertion lines show higher levels of Cyclobutane Pyrimidine Dimer (CPD) on exposure to UV-B (Lario *et al.*, 2011, 2015). We used UV-B induced thymine dimer formation as an indirect assay to monitor the activity of MSH2 in transgenic plants in comparison with WT. Pre- and post-UV-B exposure, the thymine dimers were quantified in genomic DNA of above plants by using a thymine dimer-specific monoclonal antibody (Figure S5, Figure 2). Prior to UV-B exposure, the steady-state levels of thymine dimers in control and *MSH2* silenced lines were nearly similar. However, post-6 h UV-B exposure, *MSH2* silenced lines showed a high level of dimers than the control plants. The thymine dimer level in *MSH2*-RNAi lines was higher than the control even after 24 h dark or white light incubation (Figure 2). It appears that high accumulation of thymine dimer in *MSH2* transgenic lines may have resulted from low *in-vivo* MMR activity.

MSH2-RNAi lines display abnormal pollen formation

The lines $1T_21-11$ and $2T_27-5$ with a higher reduction in *MSH2* transcript also showed phenotypic abnormalities in T_1 and T_2 generation (Figure 3a-f). The most pronounced abnormality was in the stamen morphology (Figure 3g-h). The plants in later generations too showed abnormalities in flowers, mostly in the stamens. The progeny plants of silenced lines either failed to set fruits or set fruits with very few seeds (Figure 3i-j). These seeds on germination showed a variable degree of viability. The viable seeds albeit in reduced number were propagated to the next generation.

The examination of pollens revealed that strongly silenced lines produced pollens with the deformed shape. Unlike WT pollens that were round with a vegetative and generative cell (Figure 3k-l), the pollens of silenced lines showed the deformed shape and were either empty or contained the remnant of chromosome fragments (Figure 3m-n). It is likely that pollen grains collapsed either due to lack of nourishment in deformed anthers or *MSH2*-silencing affected some other gene.

The viability of WT and *MSH2*-RNAi pollens was examined by Alexander staining. In WT, all pollens exhibited a red-purple fluorescence, strongly reflecting the high male gametophytic fertility of these plants (Figure 3o). In contrast, *MSH2*-RNAi lines such as $1T_21-11$ displayed mostly inviable pollens (Figure 3p-r). As a consequence, *MSH2*-RNAi plants set fruits bearing ca. 5% seeds. The tomato line $1T_21-11$ with the highest reduction in *MSH2* expression showed a maximum loss in the fertility with > 90% blue stained pollen grains (Figure 3p). The fruits of $1T_21-11$ line bore very few viable seeds along with underdeveloped seeds (Figure 3j).

MSH2-RNAi lines display tetraploid meiosis progression

The deformities in the anther morphology are reported to be associated with defects in male meiosis in tomato (Rick, 1948; Jeong *et al.*, 2014) and Arabidopsis (Panoli *et al.*, 2006). Considering that observed stamen abnormalities could be associated with meiosis, we examined meiosis progression in the *MSH2* silenced lines. The pollen mother cells (PMCs) were isolated at different stages of floral bud development as outlined by Brukhin *et al.* (2003) from *MSH2* silenced lines and WT. A complete meiotic progression was examined for WT which progressed normally with characteristic landmarks for different stages of the meiosis, such as zygotene (Figure 4a), pachytene (Figure 4b), diplotene and diakinesis (Figure 4c-d) followed by metaphase (Figure 4e), early and late anaphase (Figure 4f-g), early and late telophase (Figure 4h-i).

In strongly silenced *MSH2*-RNAi lines, irrespective of the stage of floral bud development, predominant short condensed chromosomes were observed in the majority of the meiocytes. These meiocytes were either arrested at the prophase I (Figure 4j-k) or completed later stages of meiosis (Figure 4l-q). The abnormal increase in the number of chromosomes in meiocytes was suggestive of either premature loss of sister chromatid cohesion or increase in ploidy leading to the formation of diploid tetrads ((Figure 4q). In these lines, very few meiocytes were observed with WT complement of chromosomes.

To quantify the chromosome number in meiotic cells of MSH2-RNAi lines, we used Centromere Protein C (CENP-C) antibody that specifically labels the centromere. Immunolabelling of WT prophase I meiocytes with CENP-C antibody showed 12 centromeres at pachytene indicating the likely presence of 12 bivalent chromosomes (Figure 5a-c). In contrast, the meiocytes of MSH2 silenced line at a stage equivalent to prophase I showed 48 centromeres (Figure 5d-f). There are two possibilities to explain the presence of 48 centromeres. First, a loss of sister chromatid cohesion in the prophase I may lead to the appearance of 48 sister chromatids. Second, an increase in ploidy to tetraploid in the pollen mother cell followed by an absence or delay in sister chromatid cohesion may result in 48 chromosomes. Analysis of meiotic stages favors the second scenario where we can observe the occurrence of diploid tetrads where the sister chromatids separate during anaphase II. Consistent with above, among the meiocytes examined, most contained 48 chromosomes as signified by CENP-C, however, in few meiocytes, the meiosis proceeded to anaphase I with segregation of 24 chromosomes to each pole (Figure 5g-i). Beyond this, other stages of meiosis I and meiosis II was rarely observed indicating a block in the meiosis of the silenced lines. However, a few cells displayed 12 centromeres indicating that silenced lines generate haploid spores albeit at highly reduced level.

We observed that only a few tetraploid cells completed meiosis II with the formation of diploid tetrads at the end of meiosis II (Figure 4k-o). The analysis of relative proportions of meiocytes revealed that in $1T_21-11$ line, 25% meiocytes displayed abnormal tetraploid chromosomes, 67% tetraploid meiocytes and 8% displayed diploid meiosis (Table 2). Similarly, the $2T_27-5$ line had 25% meiocytes with abnormal tetraploid and 75% tetraploid nuclei giving rise to diploid microspores. Consistent with this, the anthers of *MSH2* silenced lines contained largely empty shrunken pollen at the end of meiosis. Taken together these results confirm that *MSH2*-RNAi meiocytes display abnormal meiotic progression resulting in tetraploid meiocytes. The tetraploid meiocytes were observed only in $1T_21-11$ and $2T_27-5$ lines with high silencing of *MSH2* transcript. The $3T_21-4$ line with moderate *MSH2* silencing and $5T_23-3$ line with no *MSH2* silencing displayed diploid meiocytes similar to WT (Figure S6).

Analysis of T-DNA integration sites in MSH2-RNAi lines using FPNI-PCR

We also examined whether the tetraploid meiocytes or defect in cytokinesis was caused by inactivation of some important gene due to insertion of the transgene. To locate the site of transgene insertion, FPNI-PCR was performed on genomic DNA of $1T_21-11$ and $2T_27-5$ tetraploid *MSH2*-RNAi lines. The amplified PCR product flanking the right border of inserted T-DNA was cloned and sequenced. The flanking sequence of line $1T_21-11$ showed 98% homology to *Tic22* gene in chromosome No. 9 (Table S1, Figure S7a,b) and that of $2T_27-5$ showed hit in *TRM32* gene, a member of the TON1 Recruiting Motif superfamily (Table S1, Figure S7c,d). While TRM1 reportedly associates with cortical microtubules, not all TRM proteins associate with microtubules. Moreover, the sequence diversity among the TRM family members with varied transcriptional expression patterns indicates that different members may have different roles (Struk and Dhonukshe, 2014). Considering that both $1T_21-11$ and $2T_27-5$ *MSH2*-RNAi lines show tetraploidy, it is unlikely that insertional inactivation of these genes leads to tetraploidy in the silenced lines.

Root tip cells show diploid nuclei with sporadic defect in cell plate formation

In the present study, MSH2-RNAi tomato lines generated tetraploid meiocytes. However, no predominant enlarged organ size, a characteristic of increased ploidy was observed during the vegetative growth of plants. Even though the silenced lines displayed abnormal floral organs, these too showed no significant enlargement of floral structures. To ascertain whether MSH2 silencing leads to tetraploidy in the progeny of the plants, we examined somatic chromosomal count in root tips of silenced and control seedlings. The nuclei in WT root tips displayed diploid chromosome number. Interestingly the MSH2-RNAi lines 1T₂1-11 and 2T₂7-5 with tetraploid meiocytes also displayed diploid chromosomal nuclei in root tip tissues indicating that these plants were diploid. At the same time, few cells showed no intervening cell wall between the nuclei suggesting that MSH2 silencing affected the cell plate formation (Figure 6). The loss of intervening cell wall resulted in the formation of several syncytial nuclei within a cell (Figure 6e,f). In contrast, the weakly silenced MSH2-RNAi lines with diploid meiocytes exhibited root tip with uniformly sized diploid nuclei with distinct cell plate formation (Figure 6c) indicating that loss of cell plate was a feature specific to lines with a higher degree of MSH2 silencing.

To confirm whether loss of cell plate formation was related to the tetraploidy, we examined root tips of tomato tetraploid line, LA3255. The root cells of the LA3255 line showed normal tetraploid nuclei with no defect in the cell plate formation (Figure 6d). Consistent with tetraploid nature of the LA3255 line, the staining of root tip nuclei with DAPI showed larger nuclei in LA3255 line compared to diploid wild-type and *MSH2*-RNAi lines nuclei (Figure S8).

DISCUSSION

The DNA mismatch repair is the highly conserved biological pathway that maintains the fidelity of DNA replication and genetic stability. In the mammalian system, the defects in MMR lead to abnormalities in meiosis and increases predisposition to certain types of cancer (Li, 2008). Consistent with the vital role of MMR, the *MSH2* in tomato appeared recalcitrant to mutagenesis, as evident by the absence of nonsynonymous mutations in M2 lines screened by TILLING. In the mammalian system, the polymorphism in *MSH2* significantly increases the risk of breast cancer (Hsieh *et al.*, 2016). The paucity of polymorphism in *MSH2* as revealed by EcoTILLING of tomato accessions is in conformity with the indispensable nature of the gene. Though a missense and few silent mutations were identified, these had no influence on the protein function as predicted by SIFT.

In plants, the role of *MSH2* gene in MMR has been mainly inferred based on its homology with yeast and mammalian genes. In Arabidopsis, a *MSH2* T-DNA insertional mutation accumulates mutations in microsatellites and other loci indicating a role of *MSH2* in MMR (Leonard *et al.*, 2003; Hoffman *et al.*, 2004). MSH2 reportedly plays a role in repair of the UV-B induced DNA damage in the mammalian system, (Peters *et al.*, 2003; Seifert *et al.*, 2008). In plants as well, MSH2 likely plays a role in the repair of UV-B induced Cyclobutane Pyrimidine Dimer (CPD). The Arabidopsis *MSH2* mutant on UV-B exposure shows enhanced formation of CPD suggesting a role of MSH2 in the repair of UV-B damage (Lario *et al.*, 2011). Similarly, in strongly silenced tomato *MSH2*-RNAi lines, thymine dimer formation is exacerbated by exposure to UV-B. Also, *MSH2*-deficient plants show inefficient repair of thymine dimers even on exposure to visible light that stimulates photoreactivation repair. From the preceding, it is apparent that strongly silenced *MSH2*-RNAi lines have low *in vivo* MMR activity as evident by enhanced UV-B damage.

The propagation of Arabidopsis *MSH2* T-DNA insertional mutant line revealed the accumulation of a wide variety of mutations such as morphological abnormalities, reduction in seed/silique development and loss of germination efficiency (Hoffman *et al.*, 2004). Analogously, strongly silenced tomato *MSH2*-RNAi lines also show multiple abnormalities throughout the plant development, mainly during floral development with abnormal anthers. The lines with abnormal stamens showed reduced fruit set and highly reduced seed numbers in the fruits. In contrast to Arabidopsis *MSH2* mutant lines where seed set on average was reduced to only 50% (Hoffman *et al.*, 2004), in tomato silenced lines, it was less than 10%.

The reduction in seed set can result from the failure in the embryo development. Considering that the *MSH2* silenced lines developed abnormal anthers, reduced seed set may have resulted from the failure to make viable pollens in the deformed anthers. Consistent with this, Alexander staining showed nonviable pollens with blue staining in highly silenced lines. Since WT and moderately silenced lines showed viable pollens with red staining, it can be assumed that *MSH2* silencing leads to loss of pollen viability in tomato. Thus the reduced seed set in the silenced lines is in conformity with the loss of pollen viability.

The successful pollen development in plants requires operation of proper meiosis in the developing anther (Ma, 2005). In several plants defects in the anther formation are associated with defective meiosis or meiotic arrest (Walbot and Egger, 2016). In tomato, heat-induced abnormal anther development in a conditional male sterile mutant is associated with reduced pollen set and loss of pollen viability (Rick and Boynton, 1967). It is, therefore, likely that silencing of MSH2 expression affected anther development that in turn may have affected the progression of meiosis.

Consistent with this view, the *MSH2* silenced lines showed a meiotic arrest in the anthers. The majority of pollen mother cells (PMC) of *MSH2*-RNAi lines fail to complete the meiotic cycle. These PMCs initiate meiosis as evident by the onset of zygotene. However, the meiosis gets stalled at the final stage of prophase I – diakinesis, as the chromosomes are at their most condensed form. The stalling at diakinesis is seemingly related to the increase in the ploidy as CENP-C stained cells showed the presence of 48 univalent chromosomes. The stalling of the meiosis is not total, as few CENP-C stained cells also showed 24 bivalent chromosomes, which completed meiosis presumably generating diploid pollens. However, most pollens were shrunken, and only a few showed Alexander staining. It is plausible that loss of pollen viability was largely for the diploid pollens. From the preceding, it is likely that incomplete meiosis and the formation of nonviable pollens contribute to the reduced seed set in the strongly-silenced *MSH2*-RNAi lines.

The process that generates tetraploid nuclei in the PMC remains to be investigated. In several plants including tomato during microsporogenesis, the nuclei can transfer to the neighboring cell by cytomixis (Weiling, 1965; Mursalimov *et al.*, 2013). Such intercellular transfer of nuclei can influence the ploidy level of the produced pollens. However, such cytomictic transfer of nuclei most probably does not contribute to tetraploid nuclei, as in tomato cytomixis is observed only ca. 15% of PMC (Weiling, 1965). The occurrence of a large number of polyploid pollen mother cells does not favor cytomixis to be the cause of tetraploid nuclei. The probability that tetraploid meiocytes arise due to *MSH2*-silencing-induced endomitosis also seems remote. In Chrysanthemum, the fusion of two PMC forms tetraploids, albeit this process occurs at very low frequency (Kim *et al.*, 2009), and cannot account for high numbers of tetraploid meiocytes in *MSH2* silenced lines.

The root cells of *MSH2*-RNAi-silenced lines show defective cytokinesis with multiple nuclei in syncytial cells. The similar syncytial cells may also form due to defective cytokinesis during meiocytes archesporial cells development. It is likely that the fusion of nuclei in syncytial cells gives rise to the tetraploid nuclei in PMC of *MSH2*-RNAi-silenced lines. The fusion of nuclei resulting in the stochastic formation of tetraploid PMC initials has been documented in mutants defective in cell plate formation. In Arabidopsis, the mutations in sterol methyl transferase and callose synthase lead to defect in cell plate formation predominantly in the floral tissues. Both mutants also form tetraploid meiocytes and diploid gametes in both male and female sporogenesis (de Storme *et al.*, 2013). The formation of tetraploid meiocytes in *MSH2* silenced lines is analogous to tomato *pmcd1* mutant, where cytokinesis failure in meiocytes archesporial cells generates tetraploid nuclei (de Storme and Geelan, 2013a).

The *MSH2* silenced lines show the formation of the very few viable pollens and the progeny of these plants were exclusively diploid. The strongly reduced seed set in silenced lines seems to indicate that only diploid embryos survived post-fertilization. Another reason could be defective endosperm development due to unbalanced 1:1 ratio rather than normal 1:2 ratio of the paternal and maternal genome (Scott *et al.*, 1998). The absence of triploid progeny may also result from the existence of strong triploid block in members of Solanaceae including tomato (Ehlenfeldt and Ortiz 1995; Nilsson, 1950). The absence of triploid or tetraploid progeny in *MSH2*-RNAi lines is also

14

consistent with the reported absence of triploid/tetraploid plants in progeny of tomato *pmcd1* mutant that shows a cytokinesis defect and generates diploid pollens (de Storme and Geelan, 2013a)

The pre-meiotic doubling of chromosome occurs in unisexual and parthenogenetically reproducing animal species (Itono *et al.*, 2006). It is believed that in higher plants such chromosomal doubling rarely occurs (de Storme and Geelan, 2013b). In higher plants, the doubling of chromosomes reportedly occurs after the onset of meiosis via meiotic restitution (de Storme and Geelan, 2013b). Considering that doubling of chromosomes in *MSH2* silenced line occurs before the onset of meiosis, it is of a novel nature. The very high frequency of tetraploid meiocytes also hints for a likely function of *MSH2* in the regulation of chromosomal doubling via a mechanism that remains to be deciphered. Since the formation of tetraploid meiocytes in *MSH2* silenced line is quite similar to *pmcd1* mutant, the failure of cytokinesis appears to be the main contributor to this phenomenon.

In plants, MSH2 heterodimerizes with MSH3, MSH6, and MSH7. However, none of these heterodimers reportedly have any function in the regulation of meiosis. Conversely, the *MSH4* and *MSH5* mutants in Arabidopsis and rice though exhibit normal vegetative growth, show severe fertility reduction likely related to meiotic defects (Higgins *et al.*, 2004; Lu *et al.*, 2008; Luo *et al.*, 2013; Wang *et al.*, 2016). The *MSH2*-silenced lines also affected *MSH5* and *MSH7* transcript levels which may have affected the process of meiosis. The manifested tetraploid meiocytes formation and later failure in meiosis might be due to the reduction of these three transcripts. Alternately, the increased microsatellite instability caused due to low activity of the MSH2 protein may contribute to defective cytokinesis and tetraploid meiocytes formation.

Similar to *MSH2* silencing, loss of function lines in 40S ribosomal protein S6 kinases in Arabidopsis generate tetraploid PMCs and show a high degree of pollen abortion (Henriques *et al.*, 2010). Likewise, mutations in glucan synthase-like 8 and sterol methyl transferase 2 lead to defects in somatic cell plate formation and the occasional fusion of syncytial nuclei (Chen *et al.*, 2009; Guseman *et al.*, 2010). It is, therefore, likely that T-DNA insertion in a meiosis/cytokinesis related gene may have caused the tetraploidy. The examination of two silenced lines indicated insertion in *TIC22* and *TRM32* genes; both of these genes have no reported role in meiosis or cell

plate formation. Despite having an insertion in widely unrelated genes, both the silenced lines showed identical phenotype suggesting a close correlation between *MSH2* silencing and tetraploidy.

The tetraploid formation in *MSH2* silenced lines is more akin to the proposal that DNA replication or repair defects might produce signals that block cytokinesis and yield tetraploids in cancer cells (Ganem *et al.*, 2007). While it is not reported in plants, in mice embryo fibroblasts, deficiency of *MSH2* in conjunction with *P53* revealed that MSH2 plays a role in preventing polyploidization of cells (Strathdee *et al.*, 2001). In mammalian cells, *MSH2* has been identified as a target of the E2F transcription factors (Ren et al., 2002) wherein the loss of *MSH2* inhibits the interaction of *E2F* with *MSH2*, thus affecting the *E2F* signaling. It is believed that disruption of the *E2F* signaling affects the progression of cell cycle and promotes failure of cytokinesis. Since *MSH2* have a probable role in mammalian cytokinesis, it is plausible that *MSH2* may affect cytokinesis in plants in a similar fashion. Nonetheless, it remains to be determined whether *MSH2* alone or with its interacting partners plays a direct or an indirect role in regulation ploidy and cytokinesis in plants. It is also likely that *MSH2* silencing leads to wide ramification in plant cells and observed change in the ploidy level reflect its downstream effect.

In summary, our study indicates that *MSH2* silencing leads to tetraploid meiocytes formation in tomato. Additionally, *MSH2* silencing also affects cytokinesis with multiple nuclei in syncytial cells. Though the precise mechanism of *MSH2* action remains to be deciphered, these phenotypes indicate an additional role for *MSH* gene family. Our results also suggest that similar to cytomixis, the formation of syncytial cells may contribute to the tetraploid formation in higher plants. Uncovering the role of MSH2 in the regulation of cytokinesis and nuclear fusion will be of great interest in future studies and may contribute to an alternate route to the polyploid formation.

EXPERIMENTAL PROCEDURES

Screening for SNPs/mutations in *MSH2* gene

The population used for the EcoTILLING was same as described earlier in Mohan et al. (2016) and Upadhyaya et al. (2017). Genomic DNA was isolated from tomato accessions or EMS-mutagenized M₂ lines following the procedure of Sreelakshmi et al. (2010). MSH2 gene sequence was obtained from SGN (solgenomics.net, Solyc06g069230). CODDLE software (blocks.fhcrc.org/ proweb/ coddle) was used to predict the most deleterious region of the gene. The primers for SNP/mutation screening were designed using Primer3web version 4.0.0 (bioinfo.ut.ee/primer3). (Table S2). The SNP/mutation detection and haplotype assignment were carried out as described by Mohan et al. (2016). The sequence variations (Substitutions and Indels) were detected by using 'Multalin' software (http://multalin.toulouse.inra.fr/multalin/, Corpet (1988). The SNP frequency was calculated using the formula: (Total number of SNPs detected / total length of the screened fragment) X 1000 (Frerichmann et al., 2013). PARSESNP (Taylor and Greene, 2003) was used to analyze and display the variation in gene sequences, and SIFT (Sorting intolerant from tolerant; Sim et al., 2012) was used to predict the deleterious variations for protein functions.

Cloning of a fragment of MSH2 gene in pHELLSGATE 12 RNAi vector

For RNAi silencing, a 301 bp region in 4th exon of *MSH2* showing no significant homology with any other tomato gene was selected. The 301 bp region from *MSH2* cDNA was PCR amplified using primers set with the attB sequence at the 5' end (FP-AGATTCTCCAGTGATTGTTGCT; RP- CAGATCCTGTACCAAATCCCTA).. The PCR product was cloned into a pDONR201 vector with attP1 and attP2 sites using BP Clonase enzyme (*Invitrogen*). The gene fragment from the intermediate clone was transferred to the pHELLSGATE12 vector (CSIRO, Australia) by recombination between attL1/attL2 and attR1/attR2, mediated by LR Clonase enzyme (*Invitrogen*). The confirmed clone was transformed into *Agrobacterium tumefaciens* C58C1 strain.

Generation of MSH2-RNAi lines

Tomato (cv. Arka Vikas) transformation was performed as described earlier

(Sharma *et al.*, 2009) yielding ten independent transgenic lines. The phenotyping of the transgenic plants was continually carried out in successive generations. The lines with 80-90% *MSH2* transcript silencing showed anther abnormalities. Therefore these lines were manually pollinated to propagate to next generation. Nonetheless, few T_2 lines with very strong anther-defects did not survive as they produced insufficient viable pollens and failed to set fruits. The plants were carried forward till T_4 generation. The following nomenclature was adopted for progressive generations; Line No. $2T_0$ indicates T_0 line No.2, $2T_15$ indicates T_1 progeny No. 5 of Line No. 2, $2T_25-7$ indicates its T_2 progeny No. 7 of T_1 plant No.5 of T_0 Line No. 2.

Southern blot analysis

Genomic DNA (~15 µg) of T₀ transgenic lines was digested with BamHI and SalI to release a fragment size of 2.2Kb (Figure S3), while genomic DNA (~15 µg) of T₂ transgenic lines was digested EcoRI or NotI or PacI (Figure S4) at 37°C overnight, separated on a 0.8% (w/v) agarose gel, and transferred to Hybond N⁺ membranes. αP^{32} dCTP labeled 2.2 Kb BamHI, and SalI digested (Figure S3) and *NPTII* (Figure S4) fragment was synthesized by PCR and used as the probe. The primers for amplification of *NPTII* probe amplify a 554 bp fragment in the coding region of *NPTII*. Prehybridization, hybridization, washing and developing of the blot was performed according to Sambrook and Russell (2001).

RNA Isolation and Quantitative Real-Time PCR

RNA was isolated from young leaves with three independent replicates using the method as described earlier (Gupta *et al.*, 2014). Complementary DNA (cDNA) was prepared from 2 μ g of RNA using high-capacity cDNA Archive kit (Applied Biosystems, USA). Quantitative real-time PCR was also performed as described earlier by (Gupta *et al.*, 2014). Gene-specific primers were designed from the sequences obtained from the SGN database using Primer 3 software (Table S3) while real-time data analysis were same as described by (Gupta *et al.*, 2014).

Analysis of in-vivo MSH2 activity

The *in-vivo* MSH2 activity was determined by measuring the amount of thymine dimer formation on exposure to UV-B light as described by Lario *et al.* (2011, 2015). Three samples of juvenile 6th to 7th node leaves were exposed for 6 h to UV-B radiation (2 Wm⁻², TL 20 W/01-RS, Philips narrowband) in dark room at 22°C. Wild-type leaves were treated similarly albeit without UV-B radiation. One sample was frozen at the end of 6 h treatment. The second and third sample were transferred to the darkness and white light (100 μ mol m⁻² sec⁻¹) respectively for 24 h. Samples were snap frozen and stored at -80°C till DNA extraction and thymine dimer determination.

The UV-B induced thymine dimers were quantified as described by Stapleton *et al.* (1993). Six μ g of genomic DNA isolated from UV-B treated and control leaves was denatured, dot blotted onto Hybond N+ membrane (Perkin Elmer life Sciences, Inc.) and baked at 80°C. The membrane was blocked, washed and incubated with thymine dimer-specific monoclonal antibodies (Sigma-Aldrich, 1:2000 in TBS) for 3 h at room temperature with gentle agitation. The dimers were visualized by incubating with alkaline phosphatase conjugated secondary antibody (Sigma, 1:3000) and detection by nitroblue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate reaction. The bands were quantified by densitometry using Image J software. The thymine dimers were quantified using three independent biological replicates. However, in the figures, a single representative blot and the values derived from it using ImageJ software is shown.

Meiotic Chromosome Preparation

Floral buds at the appropriate stages of the meiosis were picked according to Brukhin *et al.* (2003). The buds were fixed and stored in Carnoy's solution [ethanol: chloroform: glacial acetic acid, (6:3:1)] at 4°C or -20°C. Fixed inflorescences were rinsed with Ethanol: glacial acetic acid (3:1) at defined intervals. Procedure for meiotic chromosome preparation was same as described by (Ross *et al.*, 1997). Chromosomes were counterstained with AT-specific fluorescent dye, 4'-6-diamidino-2-phenylindole (DAPI, 1 μ g/ml) (Sigma Laboratories). All images were captured using a 63X (1.4 N.A.) oil-immersion objective on a Leica SP5 confocal microscope. Images were further processed with Adobe Photoshop CS6.

Mitotic Chromosome Preparation

Roots tips (1.5-2 cm) from 5-day-old seedlings were fixed in Carnoy's solution for 24 h followed by washing with 70% (v/v) ethanol (15 min. 3X) and storage at 4°C. Procedure for mitotic chromosome preparation was same as described by (Babich *et al.*, 1997). Then the tips were observed at 100X under Nikon compound microscope (Nikon Eclipse 80i) and photographed (Nikon NIS elements Basic version 3.1). Root tip nuclei were also assessed for nuclear size in somatic tissue by fixing in ethanol:acetic acid (3:1) fixative for at least 1 h, cleared in 70% ethanol and stained with 1µg/mL 4',6diamidino-2-phenylindole (DAPI) solution.

Immunofluorescence

For immunofluorescence, synaptonemal complexes were prepared as described by Ross et al. (1997) and Chelysheva et al. (2010). 60% (v/v) acetic acid treatment was done at 45° C for 5 min with continuous stirring with a hook to remove maximum cytoplasm debris. Chromosomes were counterstained with DAPI and observed under the microscope to ensure proper slide preparation. The slides were microwaved in 10 mM citrate buffer pH 6 for 45s at 850 W for proper fixing of the chromosomes to slides. Slides were first immersed in 0.1M phosphate buffered saline (PBST): pH 7.4 containing 0.01% (v/v) TritonX-100 for 2×5 minutes. To block non-specific antibody binding, 100 μ l of blocking buffer [1% (w/v) BSA in PBS] was applied directly to the slides. 100 µl of primary CENPC (Centromere protein C) antibody (1:100) in blocking buffer [PBS + 0.1% (v/v) Triton + 1% (w/v) BSA] was applied directly to the slides, covered with parafilm and incubated overnight at 4°C in a moist chamber. The slides were washed (2×5 minutes) with PBST before adding 100 µl of the secondary antibody. Secondary antibody, goat anti-rabbit 488 (Molecular Probes; diluted 1:500) was incubated for 90 minutes at room temperature. Finally, the slides were washed for 2×5 minutes and mounted in DAPI (10 µg/ml) in Vectashield antifade mounting medium. Slides were examined using a 63X (1.4 N.A.) oil-immersion objective on a Leica SP5 confocal microscope. Images were further processed with Adobe Photoshop CS6.

Fusion primer and nested integrated PCR (FPNI-PCR)

To locate the unknown sequences flanking the insertion sites in the transgenic lines, FPNI-PCR described by Wang *et al.* (2011) with few modifications was used. The primers used for the FPNI-PCR are listed in Table S4. The cycling conditions were followed according to Wang *et al.* (2011). All PCR products were electrophoresed on a 2.0% (w/v) agarose gel and FPNI-PCR products showing sizes more than 500 bp were chosen for sequencing. The purified PCR product was sequenced using Sanger sequencing (Macrogen, South Korea). The sequences were aligned to the tomato genome sequence (Sol Genomics Network, https://solgenomics.net/) using BLAST program.

ACKNOWLEDGEMENTS

The authors wish to thank Lorinda Anderson (Colorado State University, USA) for providing CENPC antibodies. This work was supported by International Atomic Energy Agency (IAEA) (grant no. 15632/R0-4 to RS), Department of Biotechnology, India (grant no. BT/PR/ 5275/AGR/16/465/2004; BT/PR/7002/PBD/16/1009/2012) to RS and YS, and University Grants Commission (research fellowship to SS).

Authors Contributions

Conceived and designed the experiments, SS, AKP, RS, YS, MR. Performed the experiments, SS, AKP. Analyzed the data SS, AKP, RS, YS, and MR. Wrote the paper, SS, YS, RS and MR.

SUPPORTING INFORMATION

Additional SUPPORTING INFORMATION may be found in the online version of this article.

Figure S1. *In silico* prediction of most deleterious *MSH2* gene region by CODDLE and distribution of SNPs detected in *MSH2* gene by EcoTILLING.

Figure S2. Multiple sequence alignment of mRNA sequences from tomato *MSH* family. The black box indicates the sequence of *MSH2* used for making *MSH2*-RNAi construct.

Figure S3. Generation of *MSH2*-RNAi transgenic tomato lines and Southern blot of T_0 transgenic lines.

Figure S4. Southern blot of T_2 *MSH2*-RNAi lines $1T_21-11$, $2T_27-5$ and $11T_24-1$.

Figure S5. Quantification of thymine dimer levels in genomic DNA of WT and UV-B-treated *MSH2*-RNAi lines, $3T_21$ -4 and $1T_21$ -11.

Figure S6. Male meiosis in *MSH2*-RNAi tomato line No. 3T₂1-4 (a–f) and 5T₂3-3(g-l).

Figure S7. Sequence analysis of FPNI-PCR product of Line No. 1T₂1-11 and 2T₂5-5.

Figure S8. DAPI staining of root tip nuclei of WT and different *MSH2*-RNAi lines.

Table S1. The gene sequences flanking right border of T-DNA obtained using FPNI-PCR.

Table S2. List of primers used for screening for mutations and SNPs in *MSH2* gene.

Table S3. List of genes and the primers used for transcript analysis of MMR pathway.

Table S4. List of genes and the primers used for FNPI-PCR.

REFERENCES

- Acharya, S., Wilson, T., Gradia, S., Kane, M.F., Guerrette, S., Marsischky, G.T., Kolodner, R. and Fishel, R. (1996) hMSH2 forms specific mispair-binding complexes with hMSH3 and hMSH6. *Proc. Natl. Acad. Sci.USA*, 93, 13629-13634.
- Antony, E. and Hingorani, M.M. (2003) Mismatch recognition-coupled stabilization of Msh2-Msh6 in an ATP-bound state at the initiation of DNA repair. *Biochemistry*, 42, 7682-7693.
- Babich, H., Segall, M. and Fox, K. (1997) The Allium Test--A Simple, Eukaryote Genotoxicity Assay. Amer. Biol. Teacher, 59, 580-583.
- Ban, C., Junop, M. and Yang, W. (1999) Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair. *Cell*, 97, 85-97.
- Brukhin, V., Hernould, M., Gonzalez, N., Chevalier, C. and Mouras, A. (2003) Flower development schedule in tomato *Lycopersicon esculentum* cv. sweet cherry. *Sex. Plant Rep.* 15, 311-320.
- Chen, X.-Y., Liu, L., Lee, E., Han, X., Rim, Y., Chu, H., Kim, S.-W., Sack, F. and Kim, J.-Y. (2009) The Arabidopsis callose synthase gene GSL8 is required for cytokinesis and cell patterning. *Plant Physiol.* **150**, 105-113.
- Chelysheva, L., Grandont, L., Vrielynck, N., Le Guin, S., Mercier, R. and Grelon, M. (2010) An easy protocol for studying chromatin and recombination protein dynamics during *Arabidopsis thaliana* meiosis: Immunodetection of cohesins, histones and MLH1. *Cytogenet. Genome Res.* **129**, 143-153.
- Corpet, F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucl. Acids Res.* 16, 10881-10890.
- Culligan, K.M. and Hays, J.B. (2000) Arabidopsis MutS homologs—AtMSH2, AtMSH3, AtMSH6, and a novel AtMSH7—form three distinct protein heterodimers with different specificities for mismatched DNA. *Plant Cell*, **12**, 991-1002.
- De Storme, N., De Schrijver, J., Van Criekinge, W., Wewer, V., Dörmann, P. and Geelen, D. (2013) GLUCAN SYNTHASE-LIKE8 and STEROL METHYLTRANSFERASE2 are required for ploidy consistency of the sexual reproduction system in Arabidopsis. *Plant Cell*, 25, 387-403.

- **De Storme, N. and Geelen, D.** (2013a) Pre-meiotic endomitosis in the cytokinesisdefective tomato mutant *pmcd1* generates tetraploid meiocytes and diploid gametes. J. Exp. Bot. **64**, 2345-2358.
- **De Storme, N. and Geelen, D.** (2013b) Sexual polyploidization in plants–cytological mechanisms and molecular regulation. *New Phytol.*, **198**, 670-684
- **Depeiges, A., Farget, S., Degroote, F. and Picard, G.** (2005) A new transgene assay to study microsatellite instability in wild-type and mismatch-repair defective plant progenies. *Plant Science*, **168**, 939-947.
- **Dowen, J.M., Putnam, C.D. and Kolodner, R.D.** (2010) Functional studies and homology modeling of Msh2-Msh3 predict that mispair recognition involves DNA bending and strand separation. *Mol. Cell. Biol.* **30**, 3321-3328.
- Ehlenfeldt, M. and Ortiz, R. (1995) Evidence on the nature and origins of endosperm dosage requirements in Solanum and other angiosperm genera. *Sex. Plant Rep.*8, 189-196
- Emmanuel, E., Yehuda, E., Melamed-Bessudo, C., Avivi-Ragolsky, N. and Levy, A.A. (2006) The role of AtMSH2 in homologous recombination in Arabidopsis thaliana. EMBO Reports, 7, 100-105.
- Frerichmann, S.L., Kirchhoff, M., Müller, A.E., Scheidig, A.J., Jung, C. and Kopisch-Obuch, F.J. (2013) EcoTILLING in *Beta vulgaris* reveals polymorphisms in the FLC-like gene BvFL1 that are associated with annuality and winter hardiness. *BMC Plant Biol.*, **13**, 1.
- Ganem, N.J., Storchova, Z. and Pellman, D. (2007) Tetraploidy, aneuploidy and cancer. *Curr. Opin. Genet. Dev.* 17, 157-162.
- Genschel, J., Littman, S.J., Drummond, J.T. and Modrich, P. (1998) Isolation of MutSβ from human cells and comparison of the mismatch repair specificities of MutSβ and MutSα. J. Biol. Chem., 273, 19895-19901.
- Guarné, A., Ramon-Maiques, S., Wolff, E.M., Ghirlando, R., Hu, X., Miller, J.H. and Yang, W. (2004) Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair. *EMBO J.* 23, 4134-4145.
- Gupta, S.K., Sharma, S., Santisree, P., Kilambi, H.V., Appenroth, K., Sreelakshmi,
 Y. and Sharma, R. (2014) Complex and shifting interactions of phytochromes regulate fruit development in tomato. *Plant Cell Env.* 37, 1688-1702.

- Guseman, J.M., Lee, J.S., Bogenschutz, N.L., Peterson, K.M., Virata, R.E., Xie, B., Kanaoka, M.M., Hong, Z. and Torii, K.U. (2010) Dysregulation of cell-to-cell connectivity and stomatal patterning by loss-of-function mutation in Arabidopsis chorus (glucan synthase-like 8). *Development*, 137, 1731-1741.
- Henriques, R., Magyar, Z., Monardes, A., Khan, S., Zalejski, C., Orellana, J., Szabados, L., de la Torre, C., Koncz, C. and Bögre, L. (2010) Arabidopsis S6 kinase mutants display chromosome instability and altered RBR1–E2F pathway activity. *EMBO J.* 29, 2979-2993.
- Higgins, J.D., Armstrong, S.J., Franklin, F.C.H. and Jones, G.H. (2004) The Arabidopsis MutS homolog AtMSH4 functions at an early step in recombination: evidence for two classes of recombination in Arabidopsis. *Genes Dev.*, 18, 2557-2570.
- Hoffman, P.D., Leonard, J.M., Lindberg, G.E., Bollmann, S.R. and Hays, J.B. (2004) Rapid accumulation of mutations during seed-to-seed propagation of mismatch-repair-defective Arabidopsis. *Genes Dev.* 18, 2676-2685.
- Hollingsworth, N.M., Ponte, L. and Halsey, C. (1995) MSH5, a novel MutS homolog, facilitates meiotic reciprocal recombination between homologs in Saccharomyces cerevisiae but not mismatch repair. *Genes Dev.* 9, 1728-1739.
- Hsieh, Y.C., Cho, E.C., Tu, S.H., Wu, C.H., Hung, C.S., Hsieh, M.C., Su, C.T., Liu, Y.R., Lee, C.H. and Ho, Y.S. (2016) MSH2 rs2303425 Polymorphism is Associated with Early-Onset Breast Cancer in Taiwan. Ann. Surg. Onco. 1-8.
- Itono, M., Morishima, K., Fujimoto, T., Bando, E., Yamaha, E. and Arai, K. (2006) Premeiotic endomitosis produces diploid eggs in the natural clone loach, Misgurnus anguillicaudatus (Teleostei: Cobitidae). J. Exp. Zool. 305, 513-523.
- Jacobs, A.L. and Schär, P. (2012) DNA glycosylases: in DNA repair and beyond. *Chromosoma*, **121**, 1-20.
- Jeong, A.L., Lee, S., Park, J.S., Han, S., Jang, C.-Y., Lim, J.-S., Lee, M.S. and Yang, Y. (2014) Cancerous inhibitor of protein phosphatase 2A (CIP2A) protein is involved in centrosome separation through the regulation of NIMA (never in mitosis gene A)-related kinase 2 (NEK2) protein activity. J. Biol. Chem. 289, 28-40.

- Kim, J.S., Oginuma, K. and Tobe, H. (2009) Syncyte formation in the microsporangium of Chrysanthemum (Asteraceae): a pathway to infraspecific polyploidy. J. Plant Res., 122, 439-444.
- Lario, L.D., Botta, P., Casati, P. and Spampinato, C.P. (2015) Role of AtMSH7 in UV-B-induced DNA damage recognition and recombination. J. Exp. Bot. 66, 3019-3026.
- Lario, L.D., Ramirez-Parra, E., Gutierrez, C., Casati, P. and Spampinato, C.P. (2011) Regulation of plant MSH2 and MSH6 genes in the UV-B-induced DNA damage response. J. Exp. Bot. 62, 2925-2937.
- Leonard, J.M., Bollmann, S.R. and Hays, J.B. (2003) Reduction of stability of Arabidopsis genomic and transgenic DNA-repeat sequences (microsatellites) by inactivation of AtMSH2 mismatch-repair function. *Plant Physiol.* **133**, 328-338.
- Li, G.-M. (2008) Mechanisms and functions of DNA mismatch repair. *Cell Res.* 18, 85-98.
- Lloyd, A.H., Milligan, A.S., Langridge, P. and Able, J.A. (2007) TaMSH7: a cereal mismatch repair gene that affects fertility in transgenic barley (*Hordeum vulgare* L.). *BMC Plant Biol.* 7, 67.
- Lu, X., Liu, X., An, L., Zhang, W., Sun, J., Pei, H., Meng, H., Fan, Y. and Zhang,
 C. (2008) The Arabidopsis MutS homolog AtMSH5 is required for normal meiosis. *Cell Res.* 18, 589-599.
- Luo, Q., Tang, D., Wang, M., Luo, W., Zhang, L., Qin, B., Shen, Y., Wang, K., Li, Y. and Cheng, Z. (2013) The role of OsMSH5 in crossover formation during rice meiosis. *Mol. Plant*, 6, 729-742.
- Ma, H. (2005) Molecular genetic analyses of microsporogenesis and microgametogenesis in flowering plants. Ann. Rev. Plant Biol. 56, 393-434.
- Marti, T.M., Kunz, C. and Fleck, O. (2002) DNA mismatch repair and mutation avoidance pathways. J. Cell. Physiol. 191, 28-41.
- Modrich, P. (1991) Mechanisms and biological effects of mismatch repair. Ann. Rev. Genet. 25, 229-253.
- Mohan, V., Gupta, S., Thomas, S., Mickey, H., Charakana, C., Chauhan, V.S., Sharma, K., Kumar, R., Tyagi, K., Sarma, S., Gupta, S.K., Kilambi, H.V., Nongmaithem, S., Kumari, A., Gupta, P., Sreelakshmi, Y. and Sharma. R.

(2016) Tomato Fruits Show Wide Phenomic Diversity but Fruit Developmental Genes Show Low Genomic Diversity. *PloS One*, **11**, e0152907.

- Mursalimov, S.R., Sidorchuk, Y.V. and Deineko, E.V. (2013) New insights into cytomixis: specific cellular features and prevalence in higher plants. *Planta*, **238**, 415-423.
- Narine, K.A., Felton, K.E., Parker, A.A., Tron, V.A. and Andrew, S.E. (2007) Nontumor cells from an MSH2-null individual show altered cell cycle effects post-UVB. Oncology Rep. 18, 1403-1412.
- Nilsson, E. (1950) Some experiments with tetraploid tomatoes. Hereditas, 36, 181-204.
- Panoli, A.P., Ravi, M., Sebastian, J., Nishal, B., Reddy, T.V., Marimuthu, M.P., Subbiah, V., Vijaybhaskar, V. and Siddiqi, I. (2006) At MND1 is required for homologous pairing during meiosis in Arabidopsis. *BMC Mol. Biol.* 7, 1.
- Peters, A.C., Young, L.C., Maeda, T., Tron, V.A. and Andrew, S.E. (2003) Mammalian DNA mismatch repair protects cells from UVB-induced DNA damage by facilitating apoptosis and p53 activation. *DNA repair*, **2**, 427-435.
- Reardon, J.T. and Sancar, A. (2005) Nucleotide excision repair. Pro. Nucl. Acid Res. Mol. Biol. 79, 183-235.
- Reenan, R. and Kolodner, R.D. (1992) Characterization of insertion mutations in the Saccharomyces cerevisiae *MSH1* and *MSH2* genes: evidence for separate mitochondrial and nuclear functions. *Genetics*, **132**, 975-985.
- Ren, B., Cam, H., Takahashi, Y., Volkert, T., Terragni, J., Young, R.A. and Dynlacht, B.D. (2002). E2F integrates cell cycle progression with DNA repair, replication, and G2/M checkpoints. Genes Develop, 16, 245-256.
- Rick, C.M. (1948) Genetics and development of nine male-sterile tomato mutants: *Calif. Agri.* 18, 599-633.
- Rick, C.M. and Boynton, J.E. (1967) A temperature-sensitive male-sterile mutant of the tomato. *Amer. J. Bot.* 54, 601-611.
- Ross, K., Fransz, P., Armstrong, S., Vizir, I., Mulligan, B., Franklin, F. and Jones,
 G. (1997) Cytological characterization of four meiotic mutants of Arabidopsis isolated from T-DNA-transformed lines. *Chromo. Res.* 5, 551-559.
- Sachadyn, P. (2010) Conservation and diversity of MutS proteins. *Mut. Res.* 694, 20-30.

- Sambrook, J. and Russell, D.W. (2001) Molecular cloning: a laboratory manual 3rd edition. *Cold Spring Harbor Laboratory Press, USA*.
- Sandhu, A.P.S., Abdelnoor, R.V. and Mackenzie, S.A. (2007) Transgenic induction of mitochondrial rearrangements for cytoplasmic male sterility in crop plants. *Proc. Natl. Acad. Sci. USA*, **104**, 1766-1770.
- Schofield, M.J. and Hsieh, P. (2003) DNA Mismatch Repair: Molecular Mechanisms and Biological Function. Ann. Rev. Microbiol. 57, 579-608.
- Scott, R.J., Spielman, M., Bailey, J. and Dickinson, H.G. (1998) Parent-of-origin effects on seed development in Arabidopsis thaliana. *Development*, 125, 3329-3341.
- Seifert, M., Scherer, S.J., Edelmann, W., Böhm, M., Meineke, V., Löbrich, M., Tilgen, W. and Reichrath, J. (2008) The DNA-mismatch repair enzyme hMSH2 modulates UV-B-induced cell cycle arrest and apoptosis in melanoma cells. J. Invest. Dermat. 128, 203-213.
- Sharma, M.K., Solanke, A.U., Jani, D., Singh, Y. and Sharma, A.K. (2009) A simple and efficient Agrobacterium-mediated procedure for transformation of tomato. J. Biosci. 34, 423-433.
- Sim, N.-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G. and Ng, P.C. (2012) SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucl. Acids Res.* 40, W452-W457.
- Sreelakshmi, Y., Gupta, S., Bodanapu, R., Chauhan, V.S., Hanjabam, M., Thomas, S., Mohan, V., Sharma, S., Srinivasan, R. and Sharma, R. (2010) NEATTILL: A simplified procedure for nucleic acid extraction from arrayed tissue for TILLING and other high-throughput reverse genetic applications. *Plant Meth.* 6, 3.
- Srivatsan, A., Bowen, N. and Kolodner, R.D. (2014) Mispair-specific recruitment of the Mlh1-Pms1 complex identifies repair substrates of the Saccharomyces cerevisiae Msh2-Msh3 complex. J. Biol. Chem. 289, 9352-9364.
- Stapleton, A.E., Mori, T. and Walbot, V. (1993) A simple and sensitive antibodybased method to measure UV-induced DNA damage in Zea mays. Plant Mol. Biol. Rep. 11, 230-236

- Strathdee, G., Sansom, O.J., Sim, A., Clarke, A.R. and Brown, R. (2001) A role for mismatch repair in control of DNA ploidy following DNA damage. *Oncogene*, 20, 1923.
- Struk, S. and Dhonukshe, P. (2014) MAPs: cellular navigators for microtubule array orientations in Arabidopsis. *Plant Cell Rep.* **33**, 1-21.
- Takata, M., Sasaki, M.S., Sonoda, E., Morrison, C., Hashimoto, M., Utsumi, H., Yamaguchi-Iwai, Y., Shinohara, A. and Takeda, S. (1998) Homologous recombination and non-homologous end-joining pathways of DNA doublestrand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *EMBO J.* 17, 5497-5508.
- Tam, S.M., Hays, J.B. and Chetelat, R.T. (2011) Effects of suppressing the DNA mismatch repair system on homeologous recombination in tomato. *Theor. Appl. Genet.* 123, 1445-1458.
- Taylor, N.E. and Greene, E.A. (2003) PARSESNP: a tool for the analysis of nucleotide polymorphisms. *Nucl. Acids Res.* **31**, 3808-3811.
- Trouiller, B., Schaefer, D.G., Charlot, F. and Nogue, F. (2006) MSH2 is essential for the preservation of genome integrity and prevents homeologous recombination in the moss *Physcomitrella patens*. *Nucl. Acids Res.* 34, 232-242.
- Upadhyaya, P., Tyagi, K., Sarma, S., Tamboli, V., Sreelakshmi, Y. and Sharma, R. (2017) Natural variation in folate levels among tomato (*Solanum lycopersicum*) accessions. *Food Chem.* **217**, 610-619.
- Van Marcke, I. and Angenon, G. (2013) Genomic stability in Nicotiana plants upon silencing of the mismatch repair gene MSH2. *Plant Biotech. Rep.* 7, 467-480.
- Walbot, V. and Egger, R.L. (2016) Pre-Meiotic Anther Development: Cell Fate Specification and Differentiation. Ann. Rev. Plant Biol. 67, 365-395.
- Wang, C., Wang, Y., Cheng, Z., Zhao, Z., Chen, J., Sheng, P., Yu, Y., Ma, W., Duan, E. and Wu, F. (2016) The role of OsMSH4 in male and female gamete development in rice meiosis. *J. Exp. Bot.* 67, 1447-1459.
- Wang, Z., Ye, S., Li, J., Zheng, B., Bao, M. and Ning, G. (2011) Fusion primer and nested integrated PCR (FPNI-PCR): a new high-efficiency strategy for rapid chromosome walking or flanking sequence cloning. *BMC Biotech.* 11, 1.

- Weiling, F. (1965) Light and electron microscopical observation on cytomixis and its possible relation topotocytosis. *Planta*, **67**, 182-212.
- Zhao, N., Xu, X., Wamboldt, Y., Mackenzie, S.A., Yang, X., Hu, Z., Yang, J. and Zhang, M. (2016) MutS HOMOLOG1 silencing mediates ORF220 substoichiometric shifting and causes male sterility in *Brassica juncea*. J. Exp. Bot. 67, 435-444.

Table 1. List of SNPs detected in *MSH2* gene in different tomato accessions along with amino acid changes. The scores predicted by SIFT suggests that none of the amino acid changes would affect MSH2 protein activity. **Note:** A SIFT score of less than 0.05 predicts the base substitution to be intolerant.

Accession No.	No. of SNPs	Base Change	Amino acid change	SIFT Score	Nature of SNP	
EC398695	2	G1168T	A77A	-	Synonymous	
		T1365C	V143A	1.00	Nonsynonymous	
EC520075	3	A1720T	A261A	-	Synonymous	
		T1876G	A313A	-	Synonymous	
		G3193T	V639V	-	Synonymous	
EC520052	1	T1876G	A313A	-	Synonymous	
EC521078		C1312T	F125F	-	Synonymous	
	4	G1545T	G203V	0.23	Nonsynonymous	
	4	G1808A	D291N	0.83	Nonsynonymous	
		G2733T	G545V	-	Nonsynonymous	
EC528374		G1168T	A77A	-	Synonymous	
	3	T1365C	V143A	1.00	Nonsynonymous	
		T1876G	A313A	-	Synonymous	
EC34480	4	C1312T	F125F	-	Synonymous	
		G1381C	A148A	-	Synonymous	
		G1808A	D291N	0.83	Nonsynonymous	
		G3777T	A707V	-	Nonsynonymous	
EC520078	4	C1312T	F125F	-	Synonymous	
		T1339C	A134A	-	Synonymous	
		C1796G	Q287E	0.37	Nonsynonymous	
		G1808A	D291N	0.83	Nonsynonymous	

Line No	Total No. of meiocytes count	No. of Diploid meiocytes	No. of Normal Tetraploid meiocytes	No. of Abnormal Tetraploid meiocytes	No. of seeds per fruit
WT	60	60	0	0	40-50
1T ₂ 1-11	60	5	40	15	0-5
2T ₂ 7-5	60	0	45	15	2-10
3T ₂ 1-4	60	60	0	0	40-50
5T ₂ 3-3	60	60	0	0	40-50

Table 2: Number of male meiocytes in different *MSH2*-RNAi lines with reference toWT and the number of seeds per fruit in the respective lines.

Figures Legend

Figure 1. Relative transcript levels of members of *MSH* gene family in leaves of five different *MSH2*-RNAi lines. (a) *MSH1*, (b) *MSH2*, (c) *MSH3*, (d) *MSH4*, (e) *MSH5*, (f) *MSH6*, (g) *MSH7*. The transcripts levels were expressed after normalization with two internal controls, β -actin and ubiquitin. Note: line No. 1T₂1-11 shows the reduction in *MSH2*, *MSH5*, and *MSH7* transcript. The statistical significance was determined by Student's t-test. * P<0.05.

Figure 2. Determination of the *in-vivo* activity of MSH2 by quantification of thymine dimer levels using a thymine dimer-specific antibody in UV-B irradiated WT and *MSH2*-RNAi lines by dot-blot assay. (a, b) Quantification of thymine dimers in WT and *MSH2*-RNAi plants after 6 h with or without UV-B exposure (untreated). (a) Representative dot-blot. The label on top of each spot indicates the line number of *MSH2* silenced line. (b) Quantification of thymine dimers in WT and *MSH2*-RNAi lines by Image J analysis. (c, d) Quantification of thymine dimers in WT and *MSH2*-RNAi plants. Thymine dimers were quantified either immediately after 6 h exposure with UV-B, or after 24 h of white light, or dark incubation. (c) Representative dot-blot. The labels on the right of the blot indicate the line number of *MSH2* silenced line. (d) Quantification of thymine dimers in *MSH2*-RNAi lines by Image J analysis. Note: The labels on the right of the blot indicate the line number of *MSH2* silenced line. (d) quantification of thymine dimers in (b) and (d) were quantified using three independent biological replicates. The statistical significance was determined by Student's t-test. * P<0.05.

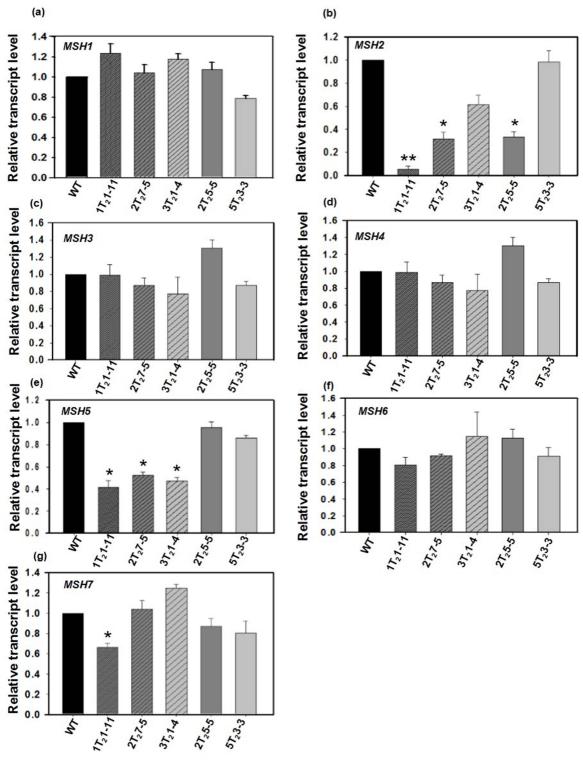
Figure 3. Abnormal flower, anther, fruit and pollen morphology manifested by *MSH2*-RNAi T₂ Plants. (a) WT flower (b-d) Abnormal floral morphology, particularly in stamens, in strongly *MSH2* silenced lines. (e-f) Flowers of partially *MSH2* silenced lines, $5T_23$ -3 and $3T_21$ -4 are nearly similar to the WT. (g) Anther cones of line WT. (h) Abnormal anther cones of line $1T_21$ -11. Note the absence of characteristic yellow coloration in $1T_21$ -11. (i-j) Vertical cut section of red ripe fruit showing highly reduced seed set in line $1T_21$ -11 (i) compared to WT (j). (k-l) WT pollen grains are bicellular with vegetative and generative cells compared to empty (m) and shriveled (n) pollen

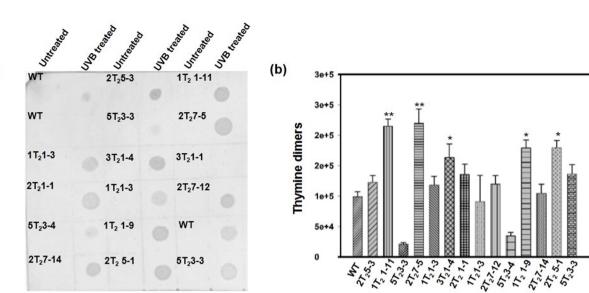
grain in line $1T_21-11$ (K-N, Scale bar, 20 µm). (o-r) Alexander staining showing viable pollen grains in WT (o), line $1T_21-11$ (p) $2T_27-5$ (q) and $2T_25-5$ (r). (l-o, Scale bar, 200 µm).

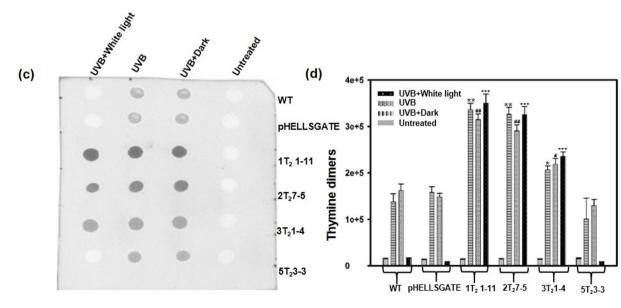
Figure 4. Male meiosis in wild-type and *MSH2*-RNAi tomato lines. Representative meiotic stages of wild-type (a–i) and *MSH2*-RNAi (j-q) lines from zygotene to telophase. In *MSH2*-RNAi meiocytes, few univalent chromosomes were observed that were assumed to be in early prophase I, (j-k). A proportion of male meiocytes of *MSH2*-RNAi lines underwent tetraploid meiosis that progressed from prophase to telophase (l-q). Note the formation of diploid tetrads (q). Scale bar, 10 μ m.

Figure 5. Immunolocalization of centromere marker protein- CENPC in nuclei of WT and *MSH2*-RNAi, $1T_21$ -11 line during prophase I. For each nuclei, the CENPC and DAPI staining along with merged signals is shown. (a-c) Zygotene nuclei with 12 centromere spots are likely indicating 12 bivalents in WT (a) with DAPI (b) and merged image (c). (d-f) Assumed zygotene nuclei with 48 centromere spots in $1T_21$ -11 line (d) are likely indicating 48 univalents along with DAPI (e) and merged image (f). (g-i) Assumed zygotene nuclei with 24 centromere spots in line $1T_21$ -11 (g) likely indicating the presence of 24 bivalents along with DAPI (h) and merged images (i). The nuclei were examined using a 63X (1.4 N.A.) oil-immersion objective on a Leica SP5 confocal microscope. Scale Bar, 10 µm.

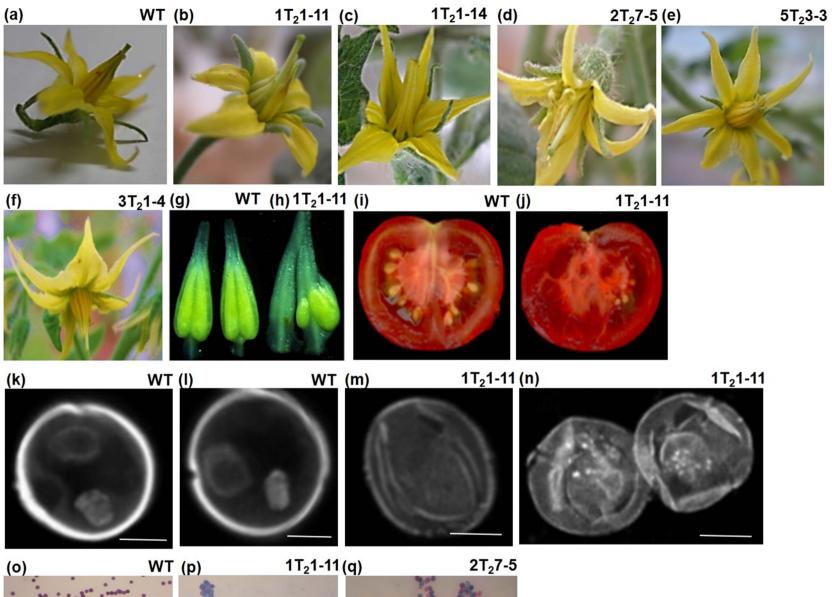
Figure 6. Mitosis in WT and T₃ *MSH2*-RNAi lines root tip cells of progenies of (a-d). WT (a), empty vector control lines (b), Progenies of *MSH2*-RNAi, $2T_21$ -1 line bearing diploid meiocytes (c), and tetraploid line LA3255 (d) showing clear separation of the diploid nucleus in mitotic cells of root tip by intervening cell plate. (e-f). Progenies of strongly silenced *MSH2*-RNAi line $2T_25$ -1 (e) and $1T_21$ -11 (f) with tetraploid meiocytes show the absence of cell plate between dividing mitotic nuclei. Scale Bar, 100 µm.

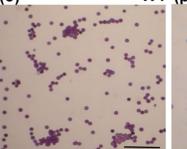


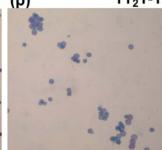


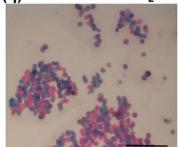


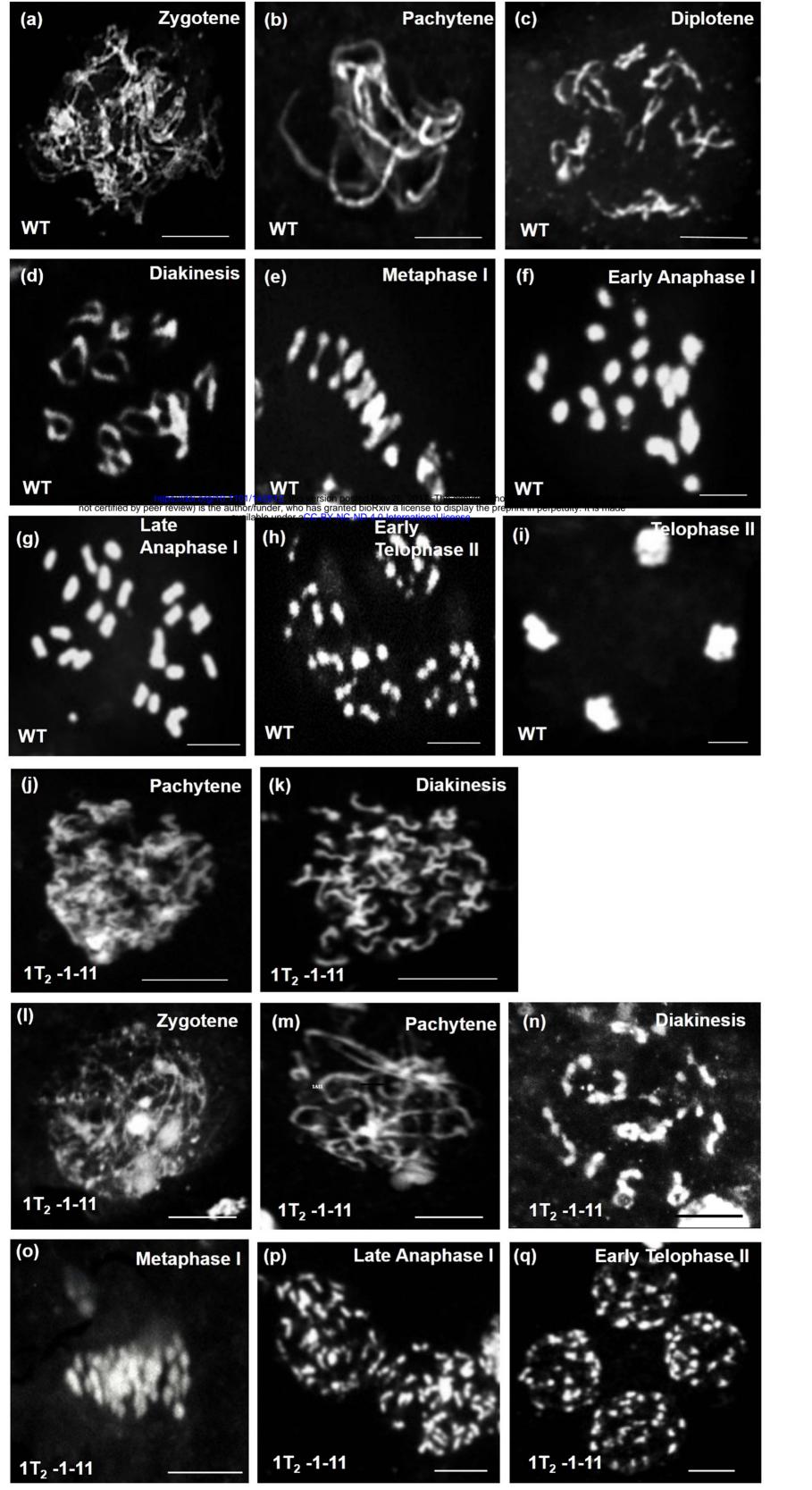
(a)

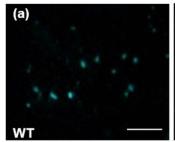


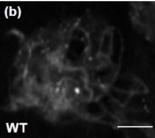


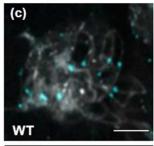


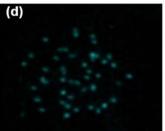




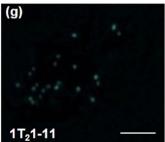




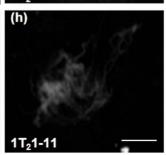


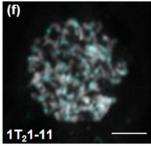


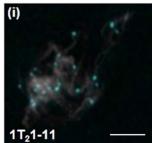


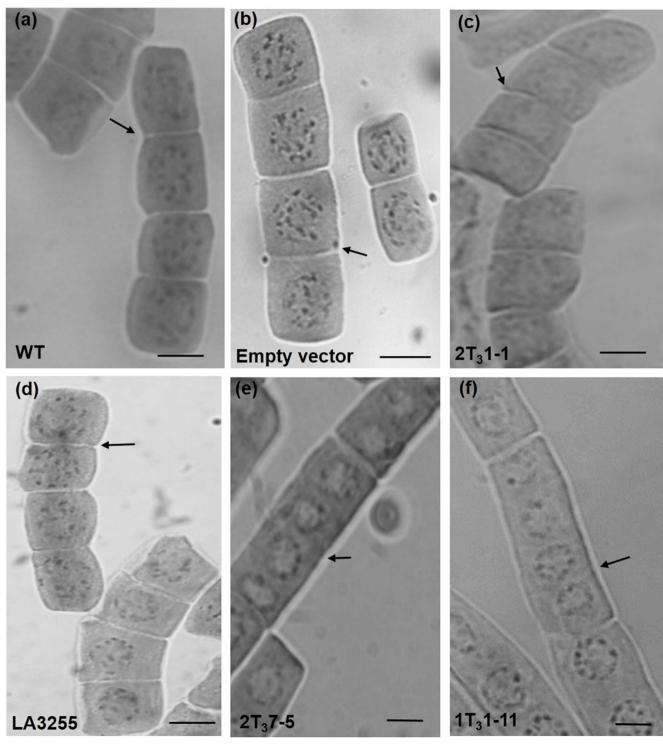


(e) 1T₂1-11 —









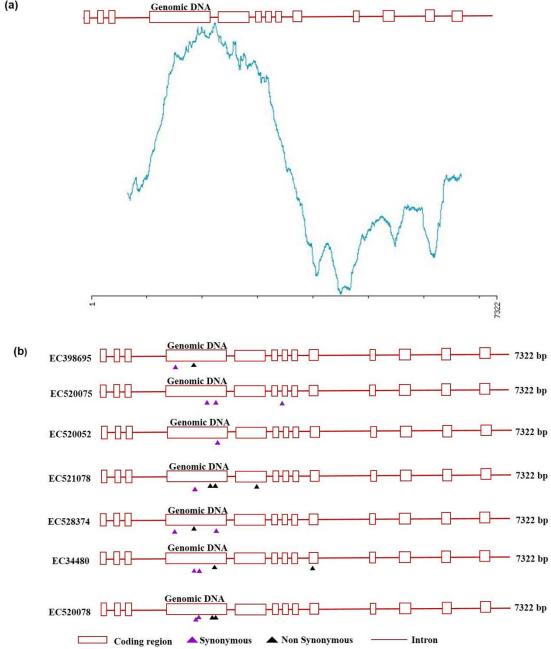


Figure S1. In silico prediction of most deleterious MSH2 gene region by CODDLE (a) and distribution of SNPs detected in MSH2 gene by EcoTILLING (b). (a) In silico prediction of most deleterious MSH2 gene region by CODDLE. The tomato MSH2 gene consists of 7322 bp with 13 exons and 12 introns. The probability curve traced in blue represents the region of the gene where mutations would most likely deleteriously affect the function of encoded protein. Based on above prediction, the segment encompassing exons 3 to 9 was chosen for EcoTILLING. (b) Distribution of SNPs detected in MSH2 gene by EcoTILLING. Red boxes denote exons interconnected with introns (solid red line). The black upright triangle indicates missense or nonsynonymous changes in the DNA sequence. The purple upright triangle indicates silent or synonymous changes. The numbers on the left and right of pictorial diagram respectively represent the accession number and total length of genomic DNA.

(a)

bioRxiv preprint doi: https://doi.org/10.1101/142612; this version posted May 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure S2. Multiple sequence alignment of mRNA sequences of tomato MSH family. Black box indicate the sequence of *MSH2* used for making *MSH2*-RNAi construct.

	î	+		+		+	+			·····			+		140	
HSH1 HSH3 HSH7 HSH6 HSH2 HSH4 HSH5	A	ATECAACA	AACCCAATA	CAGACAAGAA	-CACAATTCG	GCAAGCA-TA	TGAGTTTCAA	ACACACACAC	ACACAACGTE	ACTGT-CARAA Acaaa-cgact Acagagccara	GGACCGACC-	CTCAT	CT-TCTT	CTICITCI	TC-TC6F	CCC8
nsensus		•••••	•••••	•••••	•••••	•••••	•••••	•••••			•••••	•••••		•••••		••••
	151	160	170	180	190	200	210	220	230	240	250	260	270	280	290	3
HSH1 HSH3 HSH7 HSH6 HSH2 HSH4 HSH5	CA	TATGAGTT	CRAATTICA	ATGTACCTGA	TCCTTATCCT	GTTTTCRAGC	CCTAATTCAA	GREGTTTEEF	CCGRAGRACO	GGGTTACGGCA GARGAAGACCC TTAGTATCTC- TGTGTTAGTCC	TTCRACTTCR	ACGACCCCCT	GTACTCCACC GGAAAAGGTT	TCCCARARTI TTRARGI	IGTTGCCACCO IGATGATARAG	TCRE
nsensus	301	310	320	330	340	350	360	370	380	390	400	410	420	430	440	
HSH1 HSH3 HSH7 HSH6 HSH2 HSH4 HSH4	GTCCRC TTTCRC TGATAC	CACTTCGC		TATCTTCCC	TCCACATACT CATTTCTCCT GCGAAGTCCT TAGTCGATAR	CTGTGCCGAG CAGAA-CAAA TTGGTGCCAA AAGAGTTAAG	GAGCAG CTTTCTTCTTCTT TTCTCTCTCTTC GTTTATTGGC	-ATRCGTTGC CTTRCGATGA CCCRCTARRA CRTTGGATRA	COTGRAGGAGG TTATC-CTAC ATGCAGCGGG AATRTGGTAT	CGGAAGITTIT CAAAACCCACC CAGAAAACTTT TGAAGGITGIG	TGCCACAACG AAAACCCCCTA ACTTTCTTAC TGAAGTCTTT	GCARRAAAAA AGCTATCAAC CTCAAGAAAAC TGACAGTAGT	CCRCGTAGAT CTTCGCCGGA TCTGGTGRGC	RARARGTATT RACCCATCCF GGRCCAAR-(RTTTGGTTRF	ACTECTETTEE Geteggegat Agtatgatgat	ARAQ TRAT ARCA GGTQ
HSH5 nsensus					HLLGLLI	HCCHHH1666										
	451	460	470	480	490	500	510	520	530	540	550	560	570	580	590	6
HSH4 HSH5 Isensus	AGTGCA	IRRACC-ATO	CRAACTGAA							TT-TTCTCTAT						
HSH5 nsensus	AGTGCA	610	620													ac
HSH5 ISENSUS HSH1 HSH3 HSH7 HSH6 HSH2 HSH4 HSH5	601 GGAGATO GTT-CTF ATT-ATO TCTGAGO AAGGO GCCATTI	610 TTGCRGTT TTGCRGTT TTGATGAT CACAAGT CACAAGT TTGAAGAT TTGAAGAT	620 CRAGTCRAAR CGAAGTCACA CGAGGAAAGG ATTCTTCAAR ICACTACCA	630 ITTICCAC(ITTACAGGTAC ITACAGGTAC ATGCTGCTA ATGCTGCTACCA ITGCTCTCAC	640 FTGARGTTTTG AGATTTTTG GGAATATTGC IAGGATGTTAG TGAAATTTCG	650 650 6000000000000000000000000000000	660 AGTR	670 GGTGATT GCTGCTAGAG TGTARGA- GAGGATATGG CGTAGGGACT TTCCT	580 TTTATGAGGI TTTTGGGTAT TATCCTCTGT ATTTGGAGAT ATTATACTGC GCTTAACCC1	tt	700 GRIGCTIGTE ATGGACCACF IRATGACCACF IGATGCARGTO GATGCARGTO IGATTCCRCAF	710 TTCT ATTT GGAT TTGTTGGACC TCAT ATG	720 	730 ATGC16GTT CTAGTGTAC GCTGGGAA- ARAGTGAGT ICATATTACC	740 740 TRARTCCRTT CCRCGTTTCG 	
HSH5 nsensus HSH1 HSH3 HSH7 HSH6 HSH2 HSH4 HSH5 sensus	601 GGAGATO GTT-CTF ATT-ATO TCTGAGO AAGGO GCCATTI	610 TTGCRGTT TTGCRGTT TTGATGAT CACAAGT CACAAGT TTGAAGAT TTGAAGAT	620 CRAGTCRAAR CGAAGTCAAAA CGAGGAAAGG ATTCTTCAAA ICACTACCA	630 ITTICCAC(ITTACAGGTAC ITACAGGTAC ATGCTGCTA ATGCTGCTACCA ITGCTCTCAC	640 FTGARGTTTTG AGATTTTTG GGAATATTGC IAGGATGTTAG TGAAATTTCG	650 650 6000000000000000000000000000000	660 AGTR	670 GGTGATT GCTGCTAGAG TGTARGA- GAGGATATGG CGTAGGGACT TTCCT	580 TTTATGAGGI TTTTGGGTAT TATCCTCTGT ATTTGGAGAT ATTATACTGC GCTTAACCC1	690 CTATTGGATTC CTATTGCACAT TTATGCACAT TTATGCACACAT TTATGCACACAT CTCATGGAGGAT TACGTACTATA	700 GRIGCTIGTE ATGGACCACF IRATGACCACF IGATGCARGTO GATGCARGTO IGATTCCRCAF	710 TTCT ATTT GGAT TTGTTGGACC TCAT ATG	720 	730 ATGC16GTT CTAGTGTAC GCTGGGAA- ARAGTGAGT ICATATTACC	740 740 TRARTCCRTT CCACGTTTCG —ATGCTTCT GGGAGCAAGG ATACGACCACA ATGATCGAGA	
HSH5 Isensus HSH1 HSH3 HSH7 HSH6 HSH2 HSH4 HSH5 sensus HSH1 HSH3 HSH3 HSH1 HSH2 HSH2 HSH5 HSH5	501 1	610 TTGCRGTT TTGRAGAT SCRCARGTT TTGTAGAT SCRCARGAT TCCTCCC CCCCCCCC CCCCCCCC CCCCCCCC	620 CRAGICARAR CRAGIT-GO TRACIGA-GO TGGGGAAAGG ITCTTCRR TCACTACAR TCACTACAR TCACTACAR TCACTAR	630 ITTCCRC- TTTCCRGTA 	640 Transfirtte GGANTATICC GGANTATICC GGANTATICC GGANTATICC CTGANATICC CTGANATICC CTGANATICC CTGANATICC CTGANATICC CTGANATICC CTGANATICC CTGANATICC CTGANATICC CTGANATICC CTGANATICC CTGANATICC CTGANATICC CTGANATICC	650 CTCTGTA-G SGTRGE-G-T SGRRG-G-T SGRRG-G-T SGGCRGTTG- SGCCGTTG-G SCTTC-TGCR 800 	650 AGTA AGTA AGTA GATOCA AGTA GATOCA AGT	670 	680 TTTATGARAG TTTTAGARAG TTTTGARAG TTTTGARAG ATTTAGARAG ATTTAGARAG B30 TCACACSTAR B30 TCACACSTAR B30 TCACACSTAR B30 TCACACSTAR B30 TCACACSTAR TCACSTAR TCACACSTAR TCACACSTAR TCACACSTAR TCACACSTAR TCACACSTAR TCACACSTAR TCACACSTAR TCACSTAR	690 CTATIGGATIC TITATGCACAT TITATGCACAT TITACAGAAGAA CICATGGAGAAGAA CICATGGAGAACTATA	700 SGHIGCTTGTF ATGGGCCGCF RATGGCCGC RATGGCCGC RATGGCCGC RATGGCCGC RATGGCCGC SS0 SS0 SS0 SS0 SS0 SS0 SS0 SS	710 710 1111	220 	730 ATGC16GTT CTAGTGTAC GCT16GCAC GCT16GCAC CGT17GCAC 880 MAGGCCT TATGAGGGC TATGAGGGC ATGCCAGG	740 TRABICCATT CORGETTOR OCREGITOR OCREGITOR OCREGITOR OCREGITOR 890 CETEC-TOST NORCONET TRAGCARCE BSD CETEC-TOST NORCONET CIGLANCCA BSD CETEC-TOST CIGLANCCA BSD CETECTORET	
HSH5 ISENSUS HSH1 HSH7 HSH6 HSH2 HSH4 HSH5 HSH4 HSH3 HSH4 HSH3 HSH6 HSH2 HSH6 HSH2 HSH6 HSH2 HSH6 HSH2 HSH5 ISENSUS	501 1	610 TTGCRGTT TTGRAGAT SCRCARGTT TTGTAGAT SCRCARGAT TCCTCCC CCCCCCCC CCCCCCCC CCCCCCCC	620 CRAGICARAR CRAGIT-GO TRACIGA-GO TGGGGAAAGG ITCTTCRR TCACTACAR TCACTACAR TCACTACAR TCACTAR	630 ITTCCRC- TTTCCRGTA 	640 Transfirtte GGANTATICC GGANTATICC GGANTATICC GGANTATICC CTGANATICC CTGANATICC CTGANATICC CTGANATICC CTGANATICC CTGANATICC CTGANATICC CTGANATICC CTGANATICC CTGANATICC CTGANATICC CTGANATICC CTGANATICC CTGANATICC	650 CTCTGTA-G SGTRGE-G-T SGRRG-G-T SGRRG-G-T SGGCRGTTG- SGCCGTTG-G SCTTC-TGCR 800 	650 AGTA AGTA AGTA GATOCA AGTA	670 	680 TTTATGARAG TTTTAGARAG TTTTGARAG TTTTGARAG ATTTAGARAG ATTTAGARAG B30 TCACACSTAR B30 TCACACSTAR B30 TCACACSTAR B30 TCACACSTAR B30 TCACACSTAR TCACSTAR TCACACSTAR TCACACSTAR TCACACSTAR TCACACSTAR TCACACSTAR TCACACSTAR TCACACSTAR TCACSTAR	630 CHATTGGATTC TTATGGARAGH TTATGGARAGH TTATGGARAGH TTAGGARAGH ACTACTATA 840 ATGGATTGCTATA B40 ATGGATTGCTATA ATGGATTGCTATA CHATTGCTAT STCGARTGGARAG CHATTGCTAT STCGARTGGARAGH ACTTGCTATA	700 SGHIGCTTGTF ATGGGCCGCF RATGGCCGC RATGGCCGC RATGGCCGC RATGGCCGC RATGGCCGC SS0 SS0 SS0 SS0 SS0 SS0 SS0 SS	710 710 1111	220 	730 ATGC16GTT CTAGTGTAC GCT16GCAC GCT16GCAC CGT17GCAC 880 MAGGCCT TATGAGGGC TATGAGGGC ATGCCAGG	740 TRABICCATT CORGETTOR OCREGITOR OCREGITOR OCREGITOR OCREGITOR 890 CETEC-TOST NORCONET TRAGCARCE BSD CETEC-TOST NORCONET CIGLANCCA BSD CETEC-TOST CIGLANCCA BSD CETECTORET	IGGT GGT GGT GGT GGT GGT GGT GGT GGT GGT
HSH5 Isensus HSH1 HSH7 HSH7 HSH6 HSH6 HSH6 HSH6 HSH6 HSH6	601 GGRATHCTT GGRATHCTT ATT-ATC GGRATHCT TCTAAGA GGCATTT TTACAN GGCCATTT TTACAN GGCCATTT TTACAN GGCCATTT GGCCA GGCCATTT TTCTGAA GGCCATTT GGCCGCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCATTT GGCCA GGCCA	610 116CR61T 116AT6AT 117CAT6AT 117CAT6AT 117CAT6A 117CATCAC	620 620 620 640 640 640 640 640 640 640 64	630 1111CCRC0 1111CCRC-0 1111CCRC-0 111CCRC-0 110C1CCTRC-0 100C1CCRC-0 100C1CCRC-0 100C1CCRC-0 100C1CCRC-0 100C1CCRC-0 100C1C	640 640 640 640 640 640 640 640	650 651 652 653 655 655 655 655 655 655 655	660 6610 6610 6610 6610 6610 6610 100 6610 100 10	670 	680 TTTHGARGE TTTHGARGE TTTHGARGE TTTHGARGE TTTGGARGE 830 TCGCCCGGARGE TGGCCGGARGE TGGCCGGARGE TGGCCGGARGE TGGCCGGARGE S80 TTCCGGARGE GARGETCTTA GARGETCTTC GARGETCTTC GARGETCTTC	690 CTATTGGATTC TTATGCARA TTATGCARA TTATGCARA TTAGGARGA TAGGATCATA 840 TTGCTTCCA ATGGATTCCT TCCARGTA TTGCARGCAR TTGCARGCAR TTGTGCAC TTGTGCAC TTGTGCAC TTGTGCAC TTGTGCAC TTGTGCAC	200 2	710 710 710 710 711 711 711 716 716 716 716 716 716 716	220 720 720 7176000 7176000 7176000 7176000 7176000 7180000 7180000 7180000 7180000 71800000 718000000000000000000000000000000000000	730 ATT CHAST GALL THAT GALL ANALONG THAT CHAST CHAST CALL THAT CHAST ANALONG	740 740 740 740 741 741 741 741 741 741 741 741	
HSH6 Insensus HSH1 HSH7 HSH6 HSH7 HSH6 HSH6 HSH6 HSH6 HSH6	601 GGRATHCTT GGRATHCTT ATT-ATC GGRATHCT TCTAAGA GGCATTT TTACAN GGCCATTT TTACAN GGCCATTT TTACAN GGCCATTT GGCCA GGCCATTT TTCTGAA GGCCATTT GGCCGCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTT GGCCA GGCCATTTT GGCCATTTT GGCCATTTT GGCCATTTT GGCCATTTT GGCCATTTT GGCCATTTT GGCCATTTT GGCCATTTTT GGCCATTTTT GGCCATTTTT GGCCATTTTT GGCCATTTTT GGCCATTTTTT GGCCATTTTTT GGCCATTTTTTTTT GGCCATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	610 116CR61T 116CR61T 117CATGAT 117CATGAT 117CATGAT 117CTTCC 107CTCCC 107CTCCCC 107CTCCCC 107CTCCCC 107CTCCCC 107CTCCCC 107CTCCCC 107CTCCCC 107CTCCCC 107CTCCCC 107CTCCCC 107CTCCCCC 107CTCCCCC 107CTCCCCC 107CTCCCCC 107CTCCCCC 107CTCCCCC 107CTCCCCC 107CTCCCCC 107CTCCCCC 107CTCCCCC 107CTCCCCC 107CTCCCCC 107CTCCCCC 107CTCCCCC 107CTCCCCC 107CTCCCCC 107CTCCCCC 107CTCCCCCC 107CTCCCCCC 107CTCCCCCC 107CTCCCCCC 107CTCCCCC 107CTCCCCC 107CTCCCCCC 107CTCCCCCC 107CTCCCCCCCCCCCCCCCC 107CTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	620 620 620 640 640 640 640 640 640 640 64	630 1111CCRC0 1111CCRC-0 1111CCRC-0 111CCRC-0 110C1CCTRC-0 100C1CCRC-0 100C1CCRC-0 100C1CCRC-0 100C1CCRC-0 100C1CCRC-0 100C1C	640 640 640 640 640 640 640 640	650 651 652 653 655 655 655 655 655 655 655	660 6610 6610 6610 6610 6610 6610 100 6610 100 10	670 	680 TTTHGARGE TTTHGARGE TTTHGARGE TTTHGARGE TTTGGARGE 830 TCGCCCGGARGE TGGCCGGARGE TGGCCGGARGE TGGCCGGARGE TGGCCGGARGE S80 TTCCGGARGE GARGETCTTA GARGETCTTC GARGETCTTC GARGETCTTC	630 CIAITEGENTO CIAITEGENTO CIAITEGENTO ITTITACIONE ITTITACIONE ITTITACIONE REGISTICTO REGISTICTO INCENTION REGISTICTO ITCENTO	200 2	710 710 710 710 711 711 711 716 716 716 716 716 716 716	220 720 720 7176000 7176000 7176000 7176000 7176000 7180000 7180000 7180000 7180000 71800000 718000000000000000000000000000000000000	730 ATT CHAST GALL THAT GALL ANALONG THAT CHAST CHAST CALL THAT CHAST ANALONG	740 740 740 740 741 741 741 741 741 741 741 741	IGGT GGT GGT GGT GGT GGT GGT GGT GGT GGT

bioRxiv preprint doi: https://doi.org/10.1101/142612; this version posted May 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MSH1	1201 1210 I		+	1240	1250	1260	1270	1280	1290	1300	1310	1320	1330	1340	1350
MSH3 MSH7	GCATCGAATG	TTCGGGTGGAG	GATGTATC-A	-TCAGATCG	GTTCAGTGA		TGGCTGRGGT	GATGTCTCTC	TATGAGGGC	TGCAGGARA	TRATTIGTT	RGATGITCRA	IGARARGGARG	AGGC TGARAT	
HSH6 HSH2 HSH4	TIGGRGCCCRGGRAC ATATGATGCAATATC GTTARRGGTT	GAGATGCGGGG	TATATGAA-G TGATGGTA-A	-GGAGAGCA	ACCCCATI-	GIGGATTIC	CAGAARAGRE ARGGTRGGGE		CATCING AG	AGTIGG-CTI AGCTIGIC		ATAGGETTCT CAGTAGAACC	TGTGGTTGAG TGTTCGAGAT	CARACA CTAGTCTCTA	
MSH5 Consensus	GTGATGTTCARGTTC .tgt.caag.tc	GTGCAAGTGGG	GGCCTGTTAG	CTGT6CT	GGAAAAT	G AGCGGA	TCATAGACAC	CCTTGARCTA	ARTGARTGT	GGAGTGCATI	RETTECRAT	GHCTGCHTT	TECERERITI	CICITGACAA	ATTIC-
	1351 1360			1390	1400	1410	1420	1430	1440	1450	1460	1470	1480	1490	1500
MSH1 MSH3	TCACTGTTGTTT GCCCRARTGCRATCA	irattgCratcci	ag <mark>ggaa</mark> ta <mark>a</mark> t	GGCAATGCC	TCATTIGGC	TGCCACACAAA TGTACAAGCAT	T666CCTAA1	TGTTRGCCAT	(CTA <mark>RAACA</mark> A)	TTEGTTTEG	A-AGAGTTC	G-TGCTTC	GGAGCCTCAT	TTEGTECETT	CTCTAG
MSH7 MSH6	GTCGACAGGTTGGTA	TATCTGARA	GTG <mark>GRA</mark> TTGA BCBCBTCCCB	TCTOOCOT	TGTTCAGAA	GCTCCTAGCGC	GAGGGTACAA	RGTTGGACGG	ATGGAACAG	TEGRARACATI		RAARTCTA	GAGGETETAC	TICIGITATI ATTICATES-	CGTAGA
HSH2 HSH4 HSH5	AATGTGCAGCAGG TCTTTGCTGGTT TCAARGTTGATT		TTCAATGGAT GCTCATGAAG	CTTTTGACC	ACATGAACA	TIGATTCCACC TAGACAA-GCA	AGTGTTCAGE TCCGAGCCAT	ACTT	-GGAGATAR	TGAGCCTAT	GCACGC	TCTCTTCTC	GGCACAAA	n11ngn11	CRA
Consensus	tct.tgct6.tt	act	gatg.a.	ttt.acc	1550	1560	tg.gcaat 1570	1580	.ggaaataa. 1590	1600	1610	1620	1630	1640	1650
MSH1	ATTAGGGATCTTCTT	TTARATCCACC	A	GCCTAT	GAGATITCT	TCRGRCATTCR	AGA	GECATEC	RGACTTATE	TGAGTGTCA	ATGTTCAAT	ICCTGATTT	ACCTGT	ATTCATCIG	CRARGC
MSH3 MSH7 MSH6	CRACATEGAGATE-A AAACTAETTCATETA CTEARAECTCTCTRA		ARTGCATT TACGACAAGT CROCROSECO	GCAGCAACT	TGAGGTTTT ATTGGGCCT	GATGAATAACT GATGCAGTTCA G-TATGGTTCA	TEACGGATE	AGAGTCTGGC			ATCAGACCCT GRATEGTTCA		GGTTCARGGC GATTTGC	TTUTCHGGCA	TIGTEC
MSH2 MSH4		ATGTTATGGAG	A	GCRAAT	CAGATG-CT	AATAAAAATTT Actcgractat	TA	GRGGGGAC	TIGGICIC-I	ntgaatagaa Agggcaaat		TGCTGGAATG CTCTGAAA-G	GGTARAAGGT Acatagagac	TATTGCACAT RATTARCA	GTGGCT
MSH5 Consensus	ATGATGAATAAGT c.a.gagaa.t.t	GTGTAACC		gC.aat	GGGTAGACG	.ctaa.t	t	SEC.EE.	ttg.tttct	ag.g	CTATATIGG	.CTg.a.tg	CCTGAATCGG	CGICTIGAIA	CTATAT
	1651 1660		+	1690	1700	1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
MSH1 MSH3 MSH7	TGGTCAAGCTGCTTG GACTCATCCATTACG TGCTCTGAAAGTTTG	66ATA6RAACA	TGATAG-GTG	CT-CGTCTT	GATECRETT	TCRGAGAT	TECREARTCT	ATECHARC-T	CATCGRACT	ICTCATACTTO	Tet-Cites	RGATGGRAGE	TGCTGATG	TCACCAGTIC	GCRACC
HSH6 HSH2	TRAGACCAGTGGARA GRAACAGCCGTTACT	TRATARAGCCG RGATGTRGATG	GCTARA-TTG AGATTRACTG	CTTAGICTI IAGAIIG	GAAACTGAG GATTTAGTT	AGAGTECTE-A Carecattt-G	TECEACACACAC TEEAE-EATE	ACGAPATECE CTGCACTTCG	CTRGTGRAT	GAGTTGGTTC GAGGCAGCAT	TCTTTCTGR TGARAAGAR	ATTTTGG-GA Ittcagatat	TGRECEGCTE	RGARCARTIT RCACACARTC	GTGAGG
HSH4 HSH5 Consensus	TCGACTGCCTGGATG CTTTCTTCCTGTCC- LCL.ccLgLa	-GCTGRAGAAG	TITCGGICIC	TTTACG	GARACA-CT	GRAGICIGIAA	RAGRIATICO	CCGCR-TTCT	CARGARATT	CARCICICCA	H-GITCIATA	TCTACAAGTO	CRGACTGGT-	CAGCIT	TTCTGA
LCIII.	1801 1810	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950
MSH: MSH: MSH:	3 GGRGATACATCAT	ATAATIGTICA	AGTTCTGTCCI AGTTCTGTCCI	ACTATAGGAI GCACCACTT	AGGCCACC	RGAT-ATCCAG		CRAGAATCTT TGTTGATCCT		GCTACTGCTC	CTGAGTITAT	TGCAGTCAT	CAAGCTATC	TGATT-GCTC	
MSHI	2 GGHAHAGAGCCHG	TTTHT 6CHC6	ITGTHRHHCT	CINICHGIC	Regenterer	HTHCCHININ	CHHHHGIGI	TIGGHHCGII	HIGHIGGEC	HITIGCHCCE	CTHATCHGE	AHAGIATA		GRITCIU	LIGHGH
HSH HSH Consensus	5 AGAGTGTGTGCGC	TCTCCTGCATA	TARGCARAATI	ATTTGAAGT	TEECRITICI	GGATC-TCTGC	TAGAAGAGTI	GACGTCTTTG	GACCTGGAT	ITTATTGAGAG	GGCTGGAT	TITICATI-		TCAC	STIGATC
	1951 1960	1970	1980	1990	2000	2010	2020	2030	2040	2050	2060	2070	2080	2090	2100
MSH: MSH:	3 AACTTCAGCGGCT	TETTETEGARARE	GGCCGTI RGGACA-GAR	GTCARCAGG GTACTAR	ATCCATTTCC T-CTGCRARG	AGGAAGCAT		CARGTAGAAAAA TCCTGTTGAG	IGGCTGCGGAT	GCTTTATCT	TAGCCATAA CATCITCTAC	AGAAGATT-	-TCCTACCTAT ATGGTGCTGCT	TATTCARGE TRAGCTTTG	ATAA
MSH MSH MSH	7 CACGGGAGT 6 AACGGGAGT 2 AATGGAGT	TTTGATGGAGAGA TATGCACTCTCT GATGAT-ARTCA ATACTTCAATAA	GGACCATGATI	GTTGCTTTA GGAGTGTTA ATABGTTCA		GGGAGTCTT	CTGGACGAAT		ATTGATGTT ATTTGCGAA ACTTGRGBA	RATGAGGTC1			ATCATACGAAG GTGATGGTACT	STATATAAGGO ICAAAAATGGO	ATACCT ATA-T
MSH	5 INGCUINI	ATACTICANTA GTCTGTGAATTO	GGTAR	TTGGRGRGGG TTGGTGTCR	TGATTGATG	TGATGTCCT	TCATE	CGCGAGTTCC	TTTTGTGGC	CGGACACAGO	AGTGTTTTG	TGATGA	GCTGG	AGCAGATGG TGAACTGAG	CITCIT
Consensu:	s taa.T 2101 2110	.ttT.aat.a	2130	.T.g.gt.a	2150	2160	2170	2180	2190	2200	2210	2220	gcc 2230	2240	2250
HSH		CCCACTTGG-	ACCARCTAR	106	-GEAGATTTT	GTATECCCETE	AGCATGGAG	TGTATGGTTT	AAGGGAAAG		CRACTGTTT	GECTEGAAC	CECTEERERR	RACARATTAR	GCAACT
MSH3 MSH3	7 CARARTGGATGGA	CRECTECRCTEE	GAGAATCTIG	ACATATITG	RGRACAGTCE	ABRIGGAGATI	ICTTC866666	ATTATACACT	CRARTCRACI	-ACTGIAICE	CRECRITIE	GRARAGGAT	6CTC868TC81	GECTIECER	BCCCTT
MSH MSH MSH	2 GATGAGCAAGAGA 4 GATATG-GCTCGA		CAGATTCATA TGATACTAG	ATTIGCRCR CGRRGCT-R	RRCRRRCTGC TRCRCRGTCT	CRATGRICITE	ATCTACCTA-	TTGATRAGIC		-GRTRRRGGF	ACACAATTTU CAATAACAGO		-TETTTAGART	TRCCARGARE	GARGAR Gacata
Consensu	s gatg.gat.6.	tagtctg.	.agaa.ctt.	agg.t	aaca.a.tc.	atg.t	atc.tgaag.	.ttt.tat.	.ccat.	.galtta	caacat.	ga.a.g	t.ctag.at	.ac.aa.aA.	gct.
MSH	2251 2260	2270	2280	2290	2300	2310	2320	2330	2340	2350		2370			2400
MSH	3 CARGCAGCTTCAC	GATTCARAGEE					ind transform	T	T-BTAGATO	8000000	2360	BBBG-TCBB	0865	2390	TITIOR
HSH HSH HSH	7 CARAGATGTRGRG	GATTCAAAGGGG Atccacaaattg Araatcaatca	GAAGTATACTI Ggagtatacti Taggcttgati	AGTGTAG-C GTAGTTGAC	TGGTATCACE ARACTTGTTG	-CATTIGAGETE -CATTIGATA -ACAATGCTA	GAAGATGCAA Igagttgccto Icattatccao	TAAG TAAA TGCT	CTAGGTA-TI Cacaarggt Caatatctt	CACGAGGCAA- G-CCTU CGCAAG-CTTU	CTECTOPEC	ARAG-TCAA Caarg-tcaa Caargetgaa Taggctgct		TTOCTOCO	GACTTT TTAT-C TTCRA-T
	7 CARAGATGTAGAG 6 GTATCGTCCAGAA 2 CCAAAAGTCAGGA 4 CAGGGAARACT	ATCCACAAATTA ARAATCAATCAT	GGAGTATACTI TAGGCTTGATI	AGTGTAG-C GTAGTTGAC	TGGTATCACE ARACTTGTTG			TGCT	CACAAAGGT-	GCRAG-CTTC	GTECTAREEC GREATTEEET CREATTEEF		GGGTCTTGGAR CAGTACTAAGA TGGACGGGTTA	ATTECTARGE ARAGCARTACO ARAGCTAGCAT	TTRT-C
MSH	2 CCARARGICAGGA	ATCCACAAATTO ARAATCAATCAT TCCATTAGAGAA GGCAGCTAA GCCCAGCAAGTT GACTCTTCAGGA	GGAGTATACTI TAGGCTTGATI ACGTCAGGATI -ACTCTCACTI TCATCCAGGTI RCTATGAGTI	AGTGTAG-C GTAGTTGAC GCTGTAGCC ACATTGTTC CGTGARACA TGCGTTTGC	TGGTATCACE ARACTTGTTG GGATTAAAGG TTGARACACE TGGARACA TGAT	CATTTGATA ACAATGCTA GGCCCCAATCTA CRAGGATGGGG ATG GAG	AGAGTTGCCTO ACATTATCCAO ACCTTCTGTTO ATARAATTCAO ATCCGTTGCTO AGAGGGAGRAO	TAAA TGCT TTGAGTTTAG CAA GAG	ICACAAAGGT ICAATATCTT GARAGGAGTT ITACAAAACT GTCTTGAACT ITAGGAGGTT ITAGGAGGTT	CONTRACTOR CONTRAGONTO CONTRAGONTO CONTRAGONTO CONTRACTOR CONTRACTOR CONTRACTOR CONTRACTOR CONTRACTOR CONTRACTOR	GTOCTARGO GREATTGGA CREATTGGA CTGATATGGA GREATCGGTA ATGTGAGGAA CTGCAARGAC	CAAAGTGAA TAGGCTGCT IGCGGTTGCT ICCAGA ICAAGTCTGCI AAAAGAAAT	GGGTCTTGGAR CAGTACTAAGA TGGACGGGTTA	ATTOCTRAGGE ANAGCANTACC ARGECTAGCAT TTCGGTAGCAC CGAGTATAAGA TGCTGGATGATAT	STTRT-C TTCRA-T STGRAGC AGCTGTC SGACTGC INTRTCA
HSH: Consensu:	2 CCRARAGICAGGA 4 CAGGGAAAACTI 5 AAGAAAACCTAGA 5 caga.at.caa 2401 2410	ATCCACAAATTE ARAATCAATCAATCA TCCATTAGAAGAA GGCAGCTAA- GCCCAGCAAGTT GACTCTTCAGGA g.caag. 2420	GGAGTATACTI TAGGCTTGATI ACGTCAGGATI -ACTCTCAGTI TCATCCAGGTI ACTATGAGTI .c.tcg.T 2430	AGTGTAG-C GTAGTTGAC GCTGTAGCC ACATTGTC CGTGAAACA TGCGTTTGC .g.gtacl 2440	TGGTATCACH ARACTTGTTG GGATTARAGG TGGARACAC TGGARACAC TGGARACA TGAT		IGAGTTGCCTO CATTATCCAO ICCTTCTGTTO TAAAATTCAO STCCGTTGCTO GAGGGAGAAA (a.g.tgcao 2470	TAAA TGCT TTGAGTTTAG CAA GAA C	ICACAAAGGT CAATATCTTC AAAGGAGGTTI TACAAAACTI TACAAAACTI TAGGAGGTTI tAGGAGGTTI taaaa.To 2490	GCTCAAGCTTC CGCAAGCTTCC CAAAAAACTAC CAAAAAACTAC CAAAAAACTAC CTTTCATCATCATCATCA CCCCCCCC	GTGCTARGGG GRGATTGGGT CRGATTGGGT GRGATTGGG GRGATCGGTT ATGTGRGGGA CTGCARAGAG .tG.ta.G.z 2510	CAAAGTGAA TAGGCTGCT GCGGTTGCT ICCAG	GGGTCTTGGAR CAGTACTAAGA TGGACGGGTTA TGCACGGGTTA TGCACGGCCTCT AGATCTTAGAC AGCTAAAGAA GGATAACCTTC ,6,taa, 2530	ATTOCTANGGO INAGCANTACC ANGCCATAGCAT TAGGCTAGCAG CONGTATANGA TAGGTAGATAG TAGGAGATAG Sg. a. ag 2540	STTAT-C ITCRA-T STGAAGC AGCTGTC SGACTGC IATATCA 2550
HSH Consensu HSH HSH	2 CCHARMETCHGGH 4 CAGGGARMACTI 5 ARGANATGCTAGA 5 caga,at,ca.,a 2401 2410 1	ATCCACAGATTE ARAATCAATCAT TCCATTAGAGAA GGCAGCTAA GCCCAGCAAGTT GACTCTTCAGGA g.caag 2420 TATCTAAGAT TATTAGTGCCTT	GGAGTATACT TAGGCTAGGAT ACGTCAGGAT ACGTCAGGAT ACTATGAGTT CATCCAGGT ACTATGAGTT 2430	AGTGTAG-C GTAGTTGAC GCTGTAGCCO RCATTGTTC CGTGARACA TGCGTTTGC .g.gtacl 2440 IATCTTTGCA IAGCGTTG	TGGTATCACH ARACTTGTTG GGATTARAGO TTGARACACA TGARACA- TGAT	ICATTTGATA ACAATGCTA GGCCCCAATGCTA CCAAGGATGGGG 	IGAGTIGCCIU ICATTATCCAU ICCTTCTGTTU ITANAATICAU ITCCGTTGCTU ICAGGGAGAAA (2.a.g.tgcau 2470 AABA-ATCATT CGT-TTGCCA	TRCT TTGAGTTTAG CAR GAR GAR C	ICRCARAGG CRATATECTT RARAGGAGTT TACARARACT TACARARACT TACARARACT TACARARACT TACARARACT TACARARACT TACARARACT TACARACT 2490 GTGAGTGAGTGAAG AGTAATTTT	GRAGAGAAGA	GTGCTARAGGI CAGATITGGG CAGATITGGG CAGATICGT GAGATCGGT ATGTGAGGAT CTGCARAGGA .tG.ta.G.a 2510 AGATTGGATTAC	CRARAGTGAR TRAGGTGCTT IGCGGTTGCT ICCRGR ICCRGGTTGCT ICCRGGTTGCT ICCRGGTCTGCI ICCRGGTGCT ICCCRAC ITCCCCRAC ITTGCCCRAC	GOGTCTTGGAR GGGCCGAGCTAGG GGGCGGCCTCT AGATCTTAGAG GGCTAAAGAAT GGATAACCTTO , 5, t.a.,, 2530 AATCACCCAR GGTTCCAAGCTI	ATTECTANGGE ANGCANTACC ANGCTACATA CGASTACACA TCCGSTACACA CGASTACACA CASTACACACACACACACACACACACACACACACACACAC	GTTAT-C ITCAA-T STGAAGC AGCTGTC SGACTGC IATATCA 2550 2550 GTCAGGA CCCTGGC
HSH Consensu: HSH HSH HSH HSH	2 CCHARAGICAGA 4 CAGG-SARAACT 5 AAGAAATGCTAGA 2401 2410 1	ATCCRCRATT RARATCRATCAT RARATCRATCAT RARATCRATCAT GCCRCCATCATCAGGA GCCRGCCATCAGGA GCCCATCAGGA GCCCATCAGGA CCCCAGGA CCCCAGGA CCCAGATAGGCT GCCAGATAGGAT GCCCAGGATG AGCTCGGGATG	GGAGTATACTI TAGGCTTGAT ACGTCAGGAT ACGTCAGGAT ACTATGAGT CATCAGGT CAACAGAT CAACAGAT CAACAGAT CAACAGAT CAACAGAT TGATTGGTG GACTTATAG GACTTATAG	AGTGTAG-C GTAGTTGACC GCTGTAGCCI ACATTGTC CGTGARACA TGCGTTGCC 2440 2440 ACCTTGCGACA CARGGATG CARGGATG CCARGATA CCARGATA CCAGGATGC CCAGGA CCAGTAGC CCAGGATGC CCAGGATGC CCAGGATGC CCAGGATGC CCAGGATGC CCAGGATGC CCAGGATGC CCAGGATGC CCAGGATGC CCAGGATGC CCAGGATGC CCAGGATGC CCAGGATGC CCAGGATGC CCAGGATGC CCAGGATGC CCAGGATGC CCAGGATGC CCAGTAGC CCAGGATGC CCAGTAGC CCAGGATGC CCAGGATGC CCAGTAGC CCAGGATGC CCAGGATGC CCAGGATGC CCAGGATGC CCAGGATGC CCAGGATGC CCAGGATGC CCAGGATGC CCAGTA	TGGTATCACK ARACTTGTTG GGATTAARGG TTGAAACACC TGGAACACA TGAT	ICRITIGNIE RCRATGCTR RCRATGCRATGCGG 	GAGTTGCCTU CCTTCTGTTTCCGTT STGAGATTCAGT STGCGTGCGTGCTT GAGGGGGGAGAQ 2470 	TAR TGGR CR GR GR CR R RCTTCCATTAR RCTTTRCGT RCTTTRCGT	ICRCRARAGGI CRACTATCTI ARAGGAGAGTI ITRCRARACTI ITRCRARACTI ITRCRARACTI ITRCRARACTI ITRCRARACTI ITRCRACTGAGAGTI GTGAGTGAGTGAAGTI GTGAGTGGAGTGGG	G-CCTI GCRAG-CTTC GCCATGG-CTTC CARGAGACTAC GCTTCATGG TCCATGG TCCATGA CCCCCATG CARGAGAGAGA CTTCTACGG CARTGGGGCA CTACTGGGTCA	GTGCTARGGC GAGATTTGGG CTGATATGGG GAGATTGGG GAGATCGGT ATGTGAGAA ,tG,ta,G,: 2510 AAATTGGATT TGGTGGATTA GGTGGATTA GGCATGCTCT TGGGGATTTI	CRARGTGAR TRAGGCTGCT ICCAG	GGGTCTTGGRA TGGACGGCTTT TGCACGCCTCT TGCACGCCTCT TGCACGCCTCT GGATRACCTTC GGATRACCTTC GGATRACCTC 2530 ARTCACACAR GTTCCACAR GGTGCATCTG GGTGATCTG	ATTECTARAGEC ARAGCARTACC ARAGCTARCA CAGATARAGTARC CAGATARAGT TECGATAGA CTTGGAGATA 2540 TTTAACAART GTTGTGCARG GGARTGTCTG GGARACACAG GGARTGTCTG	GTTRT-C TTCRA-T STGRAGC SGACTGC GGACTGC IATATCA 80t.c 2550
HSH Consensus HSH HSH HSH HSH HSH HSH HSH	2 (CANARGITCAGGA 4 CAGG-GARAPACT 5 ARCARATACTAGA 2 aga,at,ca,.a 2401 2410 1 1 3 ATTCACAGARA 12 ATTCACAGARA 13 ATTCACAGARA 14 CTTCTGARAGARC 14 ATTCTTGARAGARC 15 ARAAGAGARA 15 TARAATCCTGAR	ATCCRCARTIN ARRATCRATCRA TCCRTTREGER GCCRECTRE- GCCCRECCRARGT GRCTCTTCREGE g.caag. 2420 TATCTRAGAT ATTRETEGECTT TCCTTCEGCCT TCCTAGECTRE GCCCGGGTRE GCCTCGGGTRE GCCTCGGGTRE GCCTCGGGTRE	GGAGTATACTI TAGGCTTGATI ACGTCAGGAT ACTCTCACTI ACTATGAGTI ACTATGAGTI ACTATGAGTI CAATATCCTT TAGATGAGGTI TGATTGGTGG TTCAAACAGTI TGCATTACAGGA	AGTGTAG-C GTAGTAGCC ACTGTAGCC ACATTAGCC ACATTAGCC ACATTAG	TGGTATCACC ARACTTGTTG ARACTTGTTG ARACACC TTGARACACC TGGARACRA- TGAT	ICRITTGAILE 	GENETIGCTI CETTECTA CETTECTA TECETICATI TECETICATI TECETICA SEGNEGEAGEAN (: a.g.lgcac 2470 - ABA-ATCATT CET-TECCA -GT-TECCA -GT-TECCA -GT-TECCA GTT-CATATC TIAGCTECTATE BERREGETETE	TR Provide State TGG Provide State TTGROTTING Provide State C Provide State G Provide State G Provide State Provide State Provide State <th>ICRCAAAAGGT CCRATATICTTC ARAAGAAACTI TITACAAAAACTI TITACAAAACTI 2490 GTGAGTAGAAACTI AGGAGCT AGTAATTTTT AGGGCT GATGAGATAGTG ATATGAATGTG ATATGAATGTG GTTAGTTTTG</th> <th>G-CCTI GCRAG-CTTC GCRAG-CTTC GCTTCATG6 GCTTCATG6 TTTTTATCATF ;cct. 2500 GARGARGARGAG GARGARGARGAG GARGAGTGCA CTACCG6GTTG ARATTCATTT ARATTCATTT</th> <th>GTGCTARAGG GAGATTAGG CAGATTAGG GAGATCGGTF ATGTGAGGAT CTCCAAAGA .tG.ta.G. 2510 AAATTAGATT GGTGGATAC SCTGGCATAT SCTCACACTAT SCTCACACTAT</th> <th>CRARAGIGARA TRAGECTECT ICEGEGITTECT ICERGETTECT ICERGETCEC ICERGETCEC ICERGETCEC 2520 INTECCRAC INTECCRAC INTECCRAC INTECCRAC INTECCRAC INTECRAC</th> <th>GGGTCTTGGRA CRGTRCTRAGG TGGACGGGTTT RGCACGGCTTCT RGCTCGCCCTCT RGTTCTARAGAR RGTCTARAGAR GGTCTCTGCCA GGTCCCACCAC GGTCCCACCAC GGTCCCACCAC GGTCCCACCAC GGTCCCACCAG GGTCCCACCAG GGTCCCACCAG GGTCCCACCAG GGTCCCACCAG GGTCCCCACCAG GGTCCCCACCAG GGTCCCCACCAG GGTCCCCACCAG GGTCCCCACCAG GGTCCCCCCCACCAG GGTCCCCCACCAG GGTCCCCCACCAG GGTCCCCCACCAG GGTCCCCCCCCCC</th> <th>ATTECTARAGE ANAGCARTACI ANAGCATACI ANAGCATACI ANAGCARTACI CEGETEGE CTEGETEGE CTEGETEGE CTEGEARATA 2540 TTTARCARAT GTETEGEARA GEARATECTEG GEARACCARG GEARACCARG GEARACCARG</th> <th>GTTAT-C TTCAA-T STGAAGC SGACTGC GACTGC IATATCA Sgt.c 2550 </th>	ICRCAAAAGGT CCRATATICTTC ARAAGAAACTI TITACAAAAACTI TITACAAAACTI 2490 GTGAGTAGAAACTI AGGAGCT AGTAATTTTT AGGGCT GATGAGATAGTG ATATGAATGTG ATATGAATGTG GTTAGTTTTG	G-CCTI GCRAG-CTTC GCRAG-CTTC GCTTCATG6 GCTTCATG6 TTTTTATCATF ;cct. 2500 GARGARGARGAG GARGARGARGAG GARGAGTGCA CTACCG6GTTG ARATTCATTT ARATTCATTT	GTGCTARAGG GAGATTAGG CAGATTAGG GAGATCGGTF ATGTGAGGAT CTCCAAAGA .tG.ta.G. 2510 AAATTAGATT GGTGGATAC SCTGGCATAT SCTCACACTAT SCTCACACTAT	CRARAGIGARA TRAGECTECT ICEGEGITTECT ICERGETTECT ICERGETCEC ICERGETCEC ICERGETCEC 2520 INTECCRAC INTECCRAC INTECCRAC INTECCRAC INTECCRAC INTECRAC	GGGTCTTGGRA CRGTRCTRAGG TGGACGGGTTT RGCACGGCTTCT RGCTCGCCCTCT RGTTCTARAGAR RGTCTARAGAR GGTCTCTGCCA GGTCCCACCAC GGTCCCACCAC GGTCCCACCAC GGTCCCACCAC GGTCCCACCAG GGTCCCACCAG GGTCCCACCAG GGTCCCACCAG GGTCCCACCAG GGTCCCCACCAG GGTCCCCACCAG GGTCCCCACCAG GGTCCCCACCAG GGTCCCCACCAG GGTCCCCCCCACCAG GGTCCCCCACCAG GGTCCCCCACCAG GGTCCCCCACCAG GGTCCCCCCCCCC	ATTECTARAGE ANAGCARTACI ANAGCATACI ANAGCATACI ANAGCARTACI CEGETEGE CTEGETEGE CTEGETEGE CTEGEARATA 2540 TTTARCARAT GTETEGEARA GEARATECTEG GEARACCARG GEARACCARG GEARACCARG	GTTAT-C TTCAA-T STGAAGC SGACTGC GACTGC IATATCA Sgt.c 2550
HSH Consensu HSH HSH HSH HSH HSH HSH	2 (CANARGITCAGOA 4 CAGG-GARAACT 5 AACAACAACAACAACAACAACAACAACAACAACAACAAC	ATCCRCARTIN ARRATCRATCRA TCCRTTREGER GCCRECTRE- GCCCRECCRARGT GRCTCTTCREGE g.caag. 2420 TATCTRAGAT ATTRETEGECTT TCCTTCEGCCT TCCTAGECTRE GCCCGGGTRE GCCTCGGGTRE GCCTCGGGTRE GCCTCGGGTRE	GGAGTATACTI TAGGCTTGATA ACTCCAGGAT ACTCCAGGAT ACTCCACGAGT ACTATGAGTT ACTATGAGTT CAATACCAGAT CAATACCAGAT TGATTGGTGC TTGATTGGTGC TTCAAACAGA TTCAAACAGA	AGTGTAG-C GTAGTAGCC ACTGTAGCC ACATTAGCC ACATTAGCC ACATTAG	TGGTATCACC ARACTTGTTG ARACTTGTTG ARACACC TTGARACACC TGGARACRA- TGAT	ICRITTGAILE 	GENETIGCTI CETTECTA CETTECTA TECETICATI TECETICATI TECETICA SEGNEGEAGEAN (: a.g.lgcac 2470 - ABA-ATCATT CET-TECCA -GT-TECCA -GT-TECCA -GT-TECCA GTT-CATATC TIAGCTECTATE BERREGETETE	TR Provide State TGG Provide State TTGROTTING Provide State C Provide State G Provide State G Provide State Provide State Provide State <th>ICRCAAAAGGT CCRATATICTTC ARAAGAAACTI TITACAAAAACTI TITACAAAACTI 2490 GTGAGTAGAAACTI AGGAGCT AGTAATTTTT AGGGCT GATGAGATAGTG ATATGAATGTG ATATGAATGTG GTTAGTTTTG</th> <th>G-CCTI GCRAG-CTTC GCRAG-CTTC GCTTCATG6 GCTTCATG6 TTTTTATCATF ;cct. 2500 GARGARGARGAG GARGARGARGAG GARGAGTGCA CTACCG6GTTG ARATTCATTT ARATTCATTT</th> <th>GTGCTARAGG GAGATTAGG CAGATTAGG GAGATCGGTF ATGTGAGGAT CTCCAAAGA .tG.ta.G. 2510 AAATTAGATT GGTGGATAC SCTGGCATAT SCTCACACTAT SCTCACACTAT</th> <th>CRARAGIGARA TRAGECTECT ICEGEGITTECT ICERGETTECT ICERGETCEC ICERGETCEC ICERGETCEC 2520 INTECCRAC INTECCRAC INTECCRAC INTECCRAC INTECCRAC INTECRAC</th> <th>GGGTCTTGGRA CRGTRCTRAGG TGGACGGGTTT RGCACGGCTTCT RGCTCGCCCTCT RGTTCTARAGAR RGTCTARAGAR GGTCTCTGCCA GGTCCCACCAC GGTCCCACCAC GGTCCCACCAC GGTCCCACCAC GGTCCCACCAG GGTCCCACCAG GGTCCCACCAG GGTCCCACCAG GGTCCCACCAG GGTCCCCACCAG GGTCCCCACCAG GGTCCCCACCAG GGTCCCCACCAG GGTCCCCACCAG GGTCCCCCCCACCAG GGTCCCCCACCAG GGTCCCCCACCAG GGTCCCCCACCAG GGTCCCCCCCCCC</th> <th>ATTECTARAGE ANAGCARTACI ANAGCATACI ANAGCATACI ANAGCARTACI CEGETEGE CTEGETEGE CTEGETEGE CTEGEARATA 2540 TTTARCARAT GTETEGEARA GEARATECTEG GEARACCARG GEARACCARG GEARACCARG</th> <th>GTTAT-C TTCAA-T STGAAGC SGACTGC GACTGC IATATCA Sgt.c 2550 </th>	ICRCAAAAGGT CCRATATICTTC ARAAGAAACTI TITACAAAAACTI TITACAAAACTI 2490 GTGAGTAGAAACTI AGGAGCT AGTAATTTTT AGGGCT GATGAGATAGTG ATATGAATGTG ATATGAATGTG GTTAGTTTTG	G-CCTI GCRAG-CTTC GCRAG-CTTC GCTTCATG6 GCTTCATG6 TTTTTATCATF ;cct. 2500 GARGARGARGAG GARGARGARGAG GARGAGTGCA CTACCG6GTTG ARATTCATTT ARATTCATTT	GTGCTARAGG GAGATTAGG CAGATTAGG GAGATCGGTF ATGTGAGGAT CTCCAAAGA .tG.ta.G. 2510 AAATTAGATT GGTGGATAC SCTGGCATAT SCTCACACTAT SCTCACACTAT	CRARAGIGARA TRAGECTECT ICEGEGITTECT ICERGETTECT ICERGETCEC ICERGETCEC ICERGETCEC 2520 INTECCRAC INTECCRAC INTECCRAC INTECCRAC INTECCRAC INTECRAC	GGGTCTTGGRA CRGTRCTRAGG TGGACGGGTTT RGCACGGCTTCT RGCTCGCCCTCT RGTTCTARAGAR RGTCTARAGAR GGTCTCTGCCA GGTCCCACCAC GGTCCCACCAC GGTCCCACCAC GGTCCCACCAC GGTCCCACCAG GGTCCCACCAG GGTCCCACCAG GGTCCCACCAG GGTCCCACCAG GGTCCCCACCAG GGTCCCCACCAG GGTCCCCACCAG GGTCCCCACCAG GGTCCCCACCAG GGTCCCCCCCACCAG GGTCCCCCACCAG GGTCCCCCACCAG GGTCCCCCACCAG GGTCCCCCCCCCC	ATTECTARAGE ANAGCARTACI ANAGCATACI ANAGCATACI ANAGCARTACI CEGETEGE CTEGETEGE CTEGETEGE CTEGEARATA 2540 TTTARCARAT GTETEGEARA GEARATECTEG GEARACCARG GEARACCARG GEARACCARG	GTTAT-C TTCAA-T STGAAGC SGACTGC GACTGC IATATCA Sgt.c 2550
HSH Consensus HSH HSH HSH HSH HSH HSH HSH	2 CCANARGETCAGOR 4 CAGG-GARAMACT 5 ANGARATGCTROM 2401 2410 1 CTTCTGAATTACT 33 ATTCACCAGANG 41 CTTCTGAATTACT 33 ATTCACCAGANGAACT 45 CTTCGAAGAACTGG 46 CAATGGAAGAACTGG 47 CTTCTGAAGCAGCTGG 47 CTTCTGAAGCTGG 48 CACAGAGAGAACTGG 49 CTCTCGAAGCTGG 40 CTCTGAAGCTGG 40 CTCTGAAGCTGG 40 CTCTGAAGCTGG 40 CTCTGAAGCTGG 40 CTCTGAAGCTGG 40 CTCCTGAAGCTGG 40 CTCCTGGAAGCTGG 40 CTCCTGGAAGCTGG 40 CTCCTGGACGCTGGACGTGG 40 CTCCTGGACGCTGGACGCTGGACGCTGGACGCTGGACGCTGGACGCTGGACGCTGGACGCTGGACGCTGGACGCTGGACGCTGGACGCTGGACGCCTGGACGCTGGACGCTGGACGCTGGACGCCTGGACGCCTGCCT	ATCCRCRAFT RARATCRATCA TCCRTTAGGGAR GCCRGCTRTAGGGA GCCRGCTRA GCCRGCTRAGG GCCRCCTCCRGG GCCTCTCCRGG GCCTCTCCRGG GCCTCTCCRGG GCCCTCCRGG TCCTCCCCC TCCTCCCCCCCCCC	GGRETHIGCI TREGETTGRI ACCTCLACI ACCTCLACI ACCTCLACI ACCTCLACI CANTRACA 2430 CRANTACA CANTRACA	ABTGING-C GETABITAGE GETABITAGE HEATIGITE CETABARCA LOSTABARCA SCALADA 2440 ATTEITAC HAGCHIG- LINGGHIG- LINGGHIG- LINGGHIG- LINGGAL 2500 	1661ATCACC AMACTISTIC GCATTAAAGG ITGAAAGAC TIGAAAGAC 2450 ATCTGACTGA 4250 ATCTGACTGACTGA 4250 ATCTGACTGACTGA 4250 ATCTGACTGACTGA 4250 ATCTGACTGACTGAC 4250 ATCTGACTGACTGAC 4250 ATCTGACTGACTGAC 4250 ATCTGACTGACTGAC 4250 ATCTGACTGACTGAC 4250 ATCTGACTGACTGAC 4250 ATCTGACTGACTGAC 4250 ATCTGACTGACTGAC 4250 ATCTGACTGACTGAC 4250 ATCTGACTGAC 4250 ATCTGACTGAC 4250 ATCTGACTGAC 4250 ATCTGACTGAC 4250 ATCTGACTGAC 4250 ATCTGACTGAC 4250 ATCTGACTGAC 4250 ATCTGACTGAC 4250 ATCTGACTGAC 4250 ATCTGACTGAC 4250 ATCTGACTGAC 4250 ATCTGACTGAC 4250 ATCTGACTGACTGAC 4250 ATCTGACTGACTGAC 4250 ATCTGACTGAC 4250 ATCTGACTGACTGACTGAC 4250 ATCTGACTGACTGAC 4250 ATCTGACTGACTGACTGACTGACTGAC 4250 ATCTGACTGACTGACTGACTGACTGACTGACTGACTGACT		GENERACE AND A CONTRACTOR AND A CONTRACT	110	CCRANAGET CARATECT II ANAGEAGAGTI IIICANAGEAGAGTI IIICANAGEAGAGTI IIIICAGAGGAGTI IIIICAGAGGAGTI IIIICAGAGTA 2430 GEGAGEAGAGA GAGTAGEAGA GAGTAGEAGA GAGTAGEAGA GAGTAGEAGA GAGTAGEAGA SAGGACT CEGAGEA CE	CGRGGGATGGTGATGGTGATGGTGGTGGTGGTGGTGGTGGTGG	GTGCTARAGG GAGATTGGG GAGATTGGG GAGATGGGT GAGATGGGA TGCTGARAGA 2510 2510 2510 2510 2510 2510 2510 2510	CARRETGAN TRAGETGET INFOGETGET INFOGETGET INFOGETGET INFOGETGET CEARG ANADASA - TANGAN - TTANGAN - TTANGAN - TTANGAN - TTANGAN - TANGAN -	CALL CALL CALL CALL CALL CALL CALL CALL		STTRI-C TTCRA-T STGRAGC SGCTGC SGCTGC IATATCA SCC CCCGGC ACCCAGA ACC-GC ACCCAGA ACCC-GC ACCCAGA ACCCAGA ACCCAGA CACCAGA CCCAGA
HSH Consensu: HSi HSi HSi KSi Consensu HSi KSi KSi HSi KSi HSi HSi HSi HSI HSI HSI HSI HSI	2 CCRARAMETCRAGO 4 CR6G-RAGNART 5 ANGARNTGCTR60 1 CTTCTGARTTACT 3 ATTCACCRARGAC 1 CTTCTGARTTACT 3 ATTCACCRARGACTGG 14 CTTCTGARTACTGARGARGA 14 CTTCTGARTACTGARGARGA 15 CTTCGARGACTGG 15 CTTCGARGACTGG 15 CTTCGARGACTGG 15 CTTCGARGACTGG 15 CTTCGARGACTGG 15 CTTCGARGACTGG 16 CTCCTGARTACTGARTGARTGARGACTGG 17 CTCCCTGCARGACTGG 17 CTCCCTGCATGAR 17 CTCCCTGCATGARTGARTGARTGARTGARTGARTGARTGARTGARTGA	ATCCRCRATT ARAGITCAATCAI TCCATTRAGAGA GGCCRGCTAB- GGCCRGCAB- GGCCRGCAB- GGCCTCTCAGGA g.c302. 2420 T-ATCCTAGGA T-ATCCTAGGAC TATTAGTGGCT GGCTCGGGTMB GGCTGGGGTMB GGCTCGGGTMB GGCTCGGGTMB GGCTCGGGTMB GGCTCGGGTMB GGCTCGGGTMB GGCTCGGGTMC TATTGGAGGCCG TCTAATTCTCT	GGRETHIRCT REGETTERH REGETTERH REGETTERH REGETTERH RETETERH RETETERH RETETERH REGETTERH REGTTERH R	ABTGING-C GETGINGCO GETGINGCO ACRIIGITC CETGANCO ACRIIGITC CETGANCA SCALAD ACTITIC CANGATA CANADA CA	1661ATCACC ARACITSTIC GART TARAGE TIGARRCAC 11GARRCAC 2450 ATCTGICTIS 4250 ATCTGICTIS 4250 ATCTGICTIS 4250 ATCTGICTIS 4250 ATCTGICTIS 1060ARGAR 400		GBGTIGCCI GGTIFTCCCM CCTITGCTG TITARABTICA TRANATICA SIGNEGAGNAM 2470 2470 2470 2470 2470 2470 2470 2470	119	CCCANAGET CANAGET ANAGEAGET ITREAMANE TECTAGACE ITREAMANE 2430 ECENECIA 2430 2430 ECENECIA 2430 ECEN	CGROGATICATION CARAGETTI TICARGECTIC ARMANARI LAR ARMANARI LAR ARMANARI LAR ARMANARIA CONTRACTATION CONTRACTOR CONTRACTON	GTOCTARGG GREATTGG GREATTGG CCRATTGG GREATCGE CTGATATGG GREATCGE CTGCATA ATTGG GREATCGE CTGCATA CTGCAT	CARRETGAN TRAGETGET INFAGETGET INFAGETGET INFAGETGET INFAGETGET 2520 2520 T-TCCCRAG T-TCCCCRAG T-TCCCCCRAG T-TCCCCRAG	CALL CALL CALL CALL CALL CALL CALL CALL		STTRI-C STGRAGC STGRAGC SGCTGC SGCTGC CTGRAGC CCTGGC SCTCCC CCCTGGC RCCART CCTAGGC STCAGGC STCAGGC RCCCART CTAGGC STCAGGC STCAGCC STCACGC STCACCARAT STCACCARAT STCACCARAT STCACCARAT STCACCARAT STCACCARAT STCACCARAT STCACCARAT STCACARAT STCACARAT STCACARAT STCACARAT STCACARAT STCACARAT STCACARAT STCACARAT STCACARAT STCACAR
HSH Concensus HSS HSS HSS HSS HSS Consensus HSS HSS HSS HSS HSS HSS HSS	2 CCANARAGICAGARA 4 CAGG-BARMART 5 ANGARATGCTBG 1 CACTAGARATGCTBG 1 CACTAGARATGCTBG 2401 2410 1 CACTAGARATGCTBG 2401 2410 1 CACTAGARAGICAGA 1 CACTAGARAGICAG	ATCCRORNETT ARAPTCRATCAT TCCATTRAGAGE GCCGCTGT CCCTTRAGAGE GCCGCCAGA CACCTTCAGAGE (g.cag), 2420 1ATCTARGAT ATTAGGACTAR TATTGAGGET TATTGAGGAT TATTGAGGAT TATTGAGGAT TATTGAGGAT TATTGAGGAT TATTGAGGAT TATTGAGGAT CTTARTTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTTCCT CTTCCT CTTTCCT CTTTCCT CTTCCT CTTTCCT CTTTCCT CTTCCT CTTTCCT CTTTCCT CTTCCT CTTTCCT CTTTCCT CTTCCT CTTTCCT CTTTCCT CTTCCT CTTTCCT CTTTCCT CTTCCT CTTTCCT CTTTCCT CTTCCT CTTTCCT CTTTCCT CTTCCT CTTTCCT CTTTCCT CTTCCT CTTTCCT CTTTCCT CTTCCT CTTTCCT CTTTCCT CTTCCT CTTCCT CTTTCCT CTTTCCT CTTCCT CTTTCCT CTTTCCT CTTCCT CTTTCCT C	GGentIHECT IndectTogHt RefCitCHCL RefCitCHCL CrafteGent CrafteGent Treategent Continent TreateGent TreateGent TreateGent TreateGent TreateGent Charachar Cha	AGTGING-C GETGINGCO AGTAITIGC: GETGINGCO CETGARACA TGCGITIGC: _g.gtac 2440 AGTCITIGC 2440 AGTC	1661811CACC ARACITGTIC GARITAANG TIGANARCA- TGANARCA- TGANARCA- TGANARCA- 2450 11CTGICTIG GCANNITGAN TICGANGTICTIGAGG TICGANGCAT SCANNIGANA CANARCA- 2500 1		GENETIGCET (CETTETCEN CETTETCEN TCCETTEGTI TCCETTEGTI STOCETTEGTI STOCETTEGTI SEGEGGEGNEN 2420 2420 2420 2420 2420 2420 2420 242	10	CCRARAGE CARATECT ARAGEAGET INCRARACE INCRARACE INCRARACE 1000 2490 2640 26	CGROGGATCTARTGGCTTC ADMANAGED A ADMANAGED A ADMANAGED A ADMANAGED A CONTRACTOR ADMANAGED ADMANAG	GTGCTARGG GRGATTGGG CCGATTGGG CCGATTGGG CCGATTGGG GRGATCGGT ARGTGGGGATG SGCTGGGATG SGCTGGGATG SGCTGGGATG SGCTGGGATG SGCTGGGATG SGCTGGGATG SGCTGGGATG SGCTGGGATG SGCTGGGATG SGCTGGGATG SGCTGGGATG SGCTGGGATG SGCTGGGATG SGCTGGGATG SGCTG SGCTG	CARRETGAN TRAGETGET INFAGETGET IGEGITTGET IGEGITTGET IGEGITGET IGEGITGET 2520 2520 2520 2520 2520 2520 2520 1	BGETCTTGGRE CAGIRETARGE CAGIRETARGE TGGRCGGETT TGGRCGCTCT TGGRCGCTCT TGGRCGCTCT SGATRACCTTC SGATRACCTTGCT GETCARGET RECCARGET AGEGTTGET REGGETTGE REGETTGE	1116C149666 1496C49674 1496C149624 1696C149624 1696C149624 1696C149624 1696C149624 1696C14964 1696C14964 1696C14964 1696C14964 1697C1496 1697C1497 1697C1497 1697C1497 1697C1497 1697C1497 1697C1497 1697C1497 1697C1497C1497 1697C1497 1697C1497 1697C1497C1	3TTR1-C TTCRAP-T TTCRAPT STGARGC SGCTGC SGCTGC CTCRAPT SGCTGC RECTGTC SGCTGC SGCTGC SGCTGC SGCTGC SGCTGC SGCTGC SGCTGCAGA CTCAGGA CTGAGGA ACCAARAT CTHSPSC CTHSPSC CTHSPSC GATTACTA BCTAGGA GATACT GATACT GATACT TTCAGAT GCT-ANT TTCAGAT
HSH Consensus HSJ HSJ HSJ Consensu HSJ HSJ HSJ HSJ HSJ HSJ HSJ HSJ HSJ HSJ	2 CCANARAGICAGARAGICA 4 CAGG-BARMARCT 5 ANGARATGCTBG 1 CTCTGARTACC 1 CTCTGARTACC 2401 2410 1 CTCTGARTACC 1 CTCTGARTACC 1 CTCTGARGARC 1 CTCTGARCTAR 1 CTCTGARCTAR 1 CTCTGARCTAR 1 CTCTGARCTAR 1 CTCTTGACTAR 1 CTCTTGACTAR 1 CTCTTCACTAR 1 CTCTCACTAR 1 CTCTCACTACTAR 1 CTCTTCACTAR 1 CTCTCACTAR 1 CTCTCACTAR 1 CTCTCACTAR 1 CTCTCACTAR 1 CTCTCACTACTACTAR 1 CTCTCACTAR 1 CTCTCACTAR 1 CTCTCACTAR 1 CTCTCACTAR 1 CTCTCACTAR 1 CTCCCCACTAR 1 CTCCCCCACTAR 1 CTCCCCCACTAR 1 CTCCCCCACTAR 1 CTCCCCCACTAR 1 CTCCCCACTACTACTACTACTACTACTACTACTACTACTA	ATCCRORNETT ARAPTCRATCAT TCCATTRAGAGE GCCGCTGT CCCTTRAGAGE GCCGCCAGA CACCTTCAGAGE (g.cag), 2420 1ATCTARGAT ATTAGGACTAR TATTGAGGET TATTGAGGAT TATTGAGGAT TATTGAGGAT TATTGAGGAT TATTGAGGAT TATTGAGGAT TATTGAGGAT CTTARTTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTATTGAGGAT CTTATTCCT CTTTCCT CTTCCT CTTTCCT CTTTCCT CTTCCT CTTTCCT CTTTCCT CTTCCT CTTTCCT CTTTCCT CTTCCT CTTTCCT CTTTCCT CTTCCT CTTTCCT CTTTCCT CTTCCT CTTTCCT CTTTCCT CTTCCT CTTTCCT CTTTCCT CTTCCT CTTTCCT CTTTCCT CTTCCT CTTTCCT CTTTCCT CTTCCT CTTTCCT CTTTCCT CTTCCT CTTTCCT CTTTCCT CTTCCT CTTCCT CTTTCCT CTTTCCT CTTCCT CTTTCCT CTTTCCT CTTCCT CTTTCCT C	GGentIHECT IndectTogHt RefCitCHCL RefCitCHCL CrafteGent CrafteGent Treategent Continent TreateGent TreateGent TreateGent TreateGent TreateGent Charachar Cha	AGTGING-C GETGINGCO AGTAITIGC: GETGINGCO CETGARACA TGCGITIGC: _g.gtac 2440 AGTCITIGC 2440 AGTC	1661811CACC ARACITGTIC GARITAANG TIGANARCA- TGANARCA- TGANARCA- TGANARCA- 2450 11CTGICTIG GCANNITGAN TICGANGTICTIGAGG TICGANGCAT SCANNIGANA CANARCA- 2500 1		GENETIGCET (CETTETCEN CETTETCEN TCCETTEGTI TCCETTEGTI STOCETTEGTI STOCETTEGTI SEGEGGEGNEN 2420 2420 2420 2420 2420 2420 2420 242	10	CCRARAGE CARATECT ARAGEAGET INCRARACE INCRARACE INCRARACE 1000 2490 2640 26	CGROGGATCTARTGGCTTC ADMANAGED A ADMANAGED A ADMANAGED A ADMANAGED A CONTRACTOR ADMANAGED ADMANAG	GTGCTARGG GRGATTGGG CCGATTGGG CCGATTGGG CCGATTGGG GRGATCGGT ARGTGGGGATG SGCTGGGATG SGCTGGGATG SGCTGGGATG SGCTGGGATG SGCTGGGATG SGCTGGGATG SGCTGGGATG SGCTGGGATG SGCTGGGATG SGCTGGGATG SGCTGGGATG SGCTGGGATG SGCTGGGATG SGCTGGGATG SGCTG SGCTG	CARRETGAR TRAGETGET INGGETGET IGEGETTGET IGEGETTGET IGEGETTGET IGEGETTGET 2520 2520 2520 2520 2520 2520 2520 1	BGETCTTGGRE CAGIRETARGE CAGIRETARGE TGGRCGGETT TGGRCGCTCT TGGRCGCTCT TGGRCGCTCT SGATRACCTTC SGATRACCTTGCT GETCARGET RECCARGET AGEGTTGET REGGETTGE REGETTGE	1116C149666 1496C49674 1496C149624 1696C149624 1696C149624 1696C149624 1696C149624 1696C14964 1696C14964 1696C14964 1696C14964 1697C1496 1697C1497 1697C1497 1697C1497 1697C1497 1697C1497 1697C1497 1697C1497 1697C1497C1497 1697C1497 1697C1497 1697C1497C1	3TTR1-C TTCRAP-T TTCRAPT STGARGC SGCTGC SGCTGC CTCRAPT SGCTGC SGCTGC SGCTGC SGCTGC SGCTGC SGCTGC SGCTGC SGCTGCAGA CTCAGGAC CTGAGAA SCTGAGAA CTGAGAA SCTGAGAA
HSH Consensus HSJ HSJ HSJ Consensu HSJ Consensu HSJ Consensu	2 CCANARASICAS	ятсяленаяття нарадования и порада и порада и порада и порада и порадити и по	GGentinection Record the record of the Record of the Recor	8CTG18G-C GTG6T16GC-D GCTG18GC-D RCH1G1TC CCTGABACA 12440 2440 ARTCTTTCC ABGC0TC- CCB6G8TCG-C CCB6G8TCG-C CCB6G8TCG-C CCB6G8TCG-C CCB6G8TCG-C CCB6G8TCG-C CCCCCCC 2550 	1661811CACC GRATTAAAG GRATTAAAG GRATTAAAG GRATTAAAG 1681 2450 1011 2450 1011 2450 1011 2450 1011 2450 1011 2450 1011 2450 1011 101		IGHETIGCCTI ICETTETCCH ICETTETCH ICETTETCH ICETTEGTI ICETTEGTI ICETTEGTI ICEGTEGTI ICEGTEGTI ICEGTEGTI ICEGTEGTI ICEGTETTECH ICEGTETTECH ICETTI ICEGTETTE ICETTI ICETETTECH ICETC	10	CCRANGGT CRANGT THCANAGE THCANAGE CONTRACT THCANAGE CONTRACT THCANAGE CONTRACT THCANAGE CONTRACT 2490 2490 2490 2490 2490 2490 2490 2490	CGROGOTOCO TIME CANAGE TO CARAGE TO CARAGE TO CARAGE TO CARAGE TO CONTRACT CONTRACTOR CO	GTGCTARGG GRGATTGGG CCGATTGGG CCGATTGGG GRGATCGGT GRGATCGGT SGGATCGGT 2510 2510 2510 2510 2510 2510 2510 2510	CARRETGAN TRAGETGET (CCAG - RA (CAG TIGCT (CCAG - RA 2520) 2520) 2520) 2520) 2520) 2520) 2520 2520	CREATE THE CARL AND A CONTRACT AND A		STIRI-C TTCRA-T TTCRA-T TTCRAC SERCIFIC SERCIFIC SERCIFIC 2550
HSH Consensus HSS HSS HSS Consensu HSS HSS Consensu HSS Consensu	2 CCANARGETCAGE 2 CCANARGETCAGE 3 ACAGE - SARARAT 2401 2410 1	ATCCREMENT IN AGREEMENT TCCATTREAGE SCCRECTIM	GGRETHIECT REGETTERFI REGETTERFI RECTURE RECTURE 2430 2430 CONTINUE 2430 2430 CONTINUE 2430	RETGING-C GRAFTAGLO GENETAGLO GENETAGLO CONTACT 2440 2440 CONTACT 2440 CONTACT 2440 CONTACT 2440 CONTACT 2590	166181760C 6001116081600 2001116081 2001116081 2001100 2001100 20011000 20011000 20001000 20001000 20001000 20001000 20001000 20001000 20001000 20001000 20001000 20001000 20001000 20001000 200000 200000 200000 200000 200000 200000 200000 200000 200000 200000 200000 200000 2000000		Generation Contractors Contrac	10	CCRARAGET ARAGEAGET INCARRECT INCARRECT INCARRECT INCARRECT 2490 COMPACT 2490 COMPACT 2490 COMPACT 2490 COMPACT 2490 COMPACT 2490 COMPACT 2490 COMPACT 2490 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2790 COMPACT 2700 COMPACT	CEREGENCE CARDENTE CONSCIENCE CON	GTGCTARGG GRGATTGGG CCGATTGGG CCGATTGGG GRGATCGGT GRGATCGGT ATGTGAGGAT 2510 2510 2510 2610 2610 2610 2610 2600 2600 2600 26	CIANGETGAN THRGETGET (CCAG RE (CAG TIGET (CCAG RE (CAG TIGET (CAG RE 2520 	CARLENCE CONTRACTORS CONTRACTO	TTTCCTARGG AMBCCARGT AMBCCARGAT TTCCGTARGCA Constraints Constraint	STTRI-C TTCRAPT STGRAGC SGRCTGC SGRCTGC SGRCTGC AGATSC 2550 1 GTCRGGR CCCTGGC CCCTGGC CCCTGGC CCCTGGC CCCTGGC CCTGGGC ATRACTA ACTCRAR CCRAGA STTT-CT GRTA
HSH Consensus HSS HSS HSS Consensu HSS HSS HSS Consensu HSS HSS HSS HSS HSS HSS HSS HSS HSS HS	2 CCANARGETCAGE 2 CCANARGETCAGE 3 ACAGE - SARARAT 2401 2410 1	ATCCREMENT IN AGREEMENT TCCATTREAGE SCCRECTIM	GGRETHIECT REGETTERFI REGETTERFI RECTURE RECTURE 2430 2430 CONTINUE 2430 2430 CONTINUE 2430	RETGING-C GRAFTAGLO GERITGIC: GERATAGLO CELEARCA 2440 CELEARCA CEL	166181760C 6001116081600 2001116081 2001116081 2001100 2001100 20011000 20011000 20001000 20001000 20001000 20001000 20001000 20001000 20001000 20001000 20001000 20001000 20001000 20001000 200000 200000 200000 200000 200000 200000 200000 200000 200000 200000 200000 200000 2000000		Generation Contractors Contrac	10	CCRARAGET ARAGEAGET INCARRECT INCARRECT INCARRECT INCARRECT 2490 COMPACT 2490 COMPACT 2490 COMPACT 2490 COMPACT 2490 COMPACT 2490 COMPACT 2490 COMPACT 2490 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2640 COMPACT 2790 COMPACT 2700 COMPACT	CEREGENCE CONSECUTION CONSECUT	GTGCTARGG GRGATTGGG CCGATTGGG CCGATTGGG GRGATCGGT GRGATCGGT ATGTGAGGAT 2510 2510 2510 2610 2610 2610 2610 2600 2600 2600 26	CIANGETGAN THRGETGET (CCAG RE (CAG TIGET (CCAG RE (CAG TIGET (CAG RE 2520 	CARLENCE CONTRACTORS CONTRACTO	TTTCCTARGG AMBCCARGT AMBCCARGAT TTCCGTARGCA Constraints Constraint	STTRI-C TTCRAPT STGRAGC SGRCTGC SGRCTGC SGRCTGC AGATSC 2550 1 GTCRGGR CCCTGGC CCCTGGC CCCTGGC CCCTGGC CCCTGGC CCTGGGC ATRACTA ACTCRAR CCRAGA STTT-CT GRTA
HSH Consensus HSJ HSJ HSJ KSJ Consensu HSJ HSJ KSJ Consensu HSJ HSJ HSJ HSJ HSJ HSJ HSJ HSJ HSJ HSJ	2 CCANARASICASIA 4 CR6G-RARACT 5 RAGARANTGCTN60 1 CARASIA 2 Caga.abc.ab.ab 2 Caga.abc.ab.ab 2 Caga.abc.abc.abc.abc.abc.abc.abc.abc.abc.ab	ATCCREMENT IT ADRETCRATCH ADRET	GGRETHIRCT REGETTGRIT RECTORGET	ABTGING-C GETABITGC-C GETABITGC-C GETABITGC-C CETABACA TCCCTTABCC RCHITGTC-C CETABACA 2440 2440 2440 2440 2440 2440 2440 244	166181760C 808718618 608118080 1668818 16889 176816 126811 126811 12681 12681 12681 126811 126811 126811 126811 126811 1		Generation of the second secon	TH	CICCANGG CIANTACTU ANAGGAGGTI IICCANGAGGTI IICCANGAGGTI IICCANGAGTI 2490 2540 2540 25640 25640 26640 26640 26640 26640 26640 26640 26640 26640 2750 2750 2750 2750 2750 2750 2750 275		GTGCTARGG GRGATTGGG CCGATTGGG CCGATTGGG CCGATTGGG GRGATCGGT CCGATTGGG GRGATCGGT CCGATTGGG SGGTCGGTC SGGTGGTTG CCGCATGG CCGATGG CCGATG CCGATGG CCGATG	CHARGTGAN THRGGTGCT INGGTGCT ICCAG INGGTGCT ICCAG INGGTGCT INGGTGCT INGGTGCT INGGTG INGGTGGTC INGGTGGTC INGGTGGTC INGGTGGTC ING ING ING ING ING ING ING ING ING ING	GBGTCTTGGRE CRGTRGTTAGGE CRGTRGTTAGGE TGGRCGGGTT TGGRCGGCTTG BGGTCTTGGRE BGGTCTGGRE GGGTAGGCTTG GGTCGRGTGCTTG GGTCGRGTGCTTG GGTCGRGTGCTGG GGTCGRGTGCTGG GGTCGRGTGCTGG GGTCGRGTGCTGG GGTGGTGCTGGT GGTGGTGCTGGT GGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	TTOCTARGG AND/CARTAC AND/CARTAC AND/CARTAC AND/CARTACA	STTRI-C TTCRA-T TTCRA-T STGRAGC BCCTGTC SGRCTGC SGRCTGC AGCA SGRCTGC AGCA SGRCTCC AGCA SGRCTCC AGCA AGCA AGCA AGCA AGCA AGCA AGCA
HSHE Consensus HSS HSS HSS HSS Consensu HSS HSS Consensus HSS HSS HSS HSS HSS HSS HSS HSS HSS HS	2 CCANARASICASION 4 CR6G-RARASIC 5 RAGARATACTRA 2401 2410 1 TTCTGRATTACC 1 ATTORCAGA 2401 2410 1 TTCTGRATTACC 1 ATTORCAGA 1 ATTORCAGA 2401 2410 1 TTCTGRARGAC 1 ATTORCAGA 1 ATTORCAGA 2401 2410 1 ATTORCAGA 1 ATTORCAGA 1 ATTORCAGA 2401 2410 1 ATTORCAGA 2551 2550 1 ATTORCAGA 2551 2550 1 ATTORCAGA 2551 2550 1 ATTORCAGACA 2551 2550 1 ATTACAGA 2551 2550 1 ATTACAGACA 2551 2550 1 A	HICCROMMITT AMARTCARTCAR TCCATTRAGAGE GCCGCCTAR GCCGCCARGET GCCGCCARCT GCCGCCARCT CCCCARCARCT CCCCARCARCT CCCCARCARCT CCCCARCARCT CCCCARCARCT CCCCCARCARCT CCCCCARCARCT CCCCCARCARCT CCCCCARCARCT CCCCARCT CCCCARCARCT CCCCCARCT CCCCCCCARCT CCCCARCT	GGRETHIRCT REGETTGRIT RECTORGET	BCTGING-C GTAGTIGGLO GCTGINGCD BCTGINGCD BCTGINGCD BCTGINGCD CCTGABACA CCTGABACA CCTGCA 2440 CTCCGTIGC 2440 CTCCGTIGC CSC CSC CSC CSC CSC CSC CSC CSC CSC C	166181CAC 6081TAAAG 6081TAAAG 16881CAC 16881CAC 16881CAC 16881CAC 16881CAC 16881CAC 16881CAC 16881CAC 169		GBGTIGCT CHTHTCCAR LCTITCTGTI TICCATTGTI TICCATTGTI STOCATTGTI STOCATTGTI STOCATTGTI STOCATTGTI STOCATTGTI ARB-ATCATT CIT-CITAC ARB-ATCATT CIT-CITAC CIT-CITACA CIT-CITACA CIT-CITACA CIT-CITACA CIT-CITACA CIT-CITACA CIT-CITACA CITCCATTG STOCATTG S	118	CICCANGG CICANGTACT ANAGEAGG TI IICCANGAG TI IICCANGAG TI IICCANGAG TI IICCANGAG TI IICCANGAG TI CANGAG TI		GTGCTARGG GRGATTGGG CCGATTGGG CCGATTGGG CCGATTGGG GRGATCGGT CCGATATGG GRGATCGGT CCGATAGGA 2510 2510 2510 2510 2510 2510 2510 2510	CHARGTGAR THRGGTGCT GCGGTTGCT GCGGTTGCT GCGGTTGCT CCAGS	GGETCTTGGRE CRETRETTAGE CRETRETTAGE TGEACGETTT TGEACGETTG BERTCTTGGRE BERTCTTGGRE SCHARCCTET GERTRECTTG GERTRECTTG TGEACGETTG GERTRECTTG GERTRECTTG TGECONCEG GGETCANTEGT GGETCANTEGT TGECONCEG GGETCANTEGT GGETCANTEGT CALL 2680 THERTAGE THEGTGANGE ANDEC THEGTGANGE GERGTTG	TTICCTARGG AND/CARTINC AND/CARTINC AND/CARTINC AND/CARTINCA CONTINUES CONTIN	3TTR1-C TTCRA-T TTCRA-T TTCRA-T TTCRA-T TTCRA-T STGRAGC BCCTGC SGRCTCC SGRCTCC 2550
HSH Consensus HSS HSS HSS Consensus HSS HSS HSS HSS HSS HSS HSS HSS HSS HS	2 CCANARGETCARGO 4 CRGG-RARMACT 5 RAGARATACTRA 2401 2410 1 CTTCGARTTACT 3 RATCACCAGANA 2401 2410 1 CTTCGARTTACT 3 RATCACCAGANA 4 CTTCGARTACTAC 4 CTTCGARTACTAC 4 CTTCGARTACTAC 4 CTTCTGARTACTAC 4 CTTCTGARTACTAC 4 CTTTGARGAGANA 4 CTTTGATGARGANA 4 CTTTGATGARGANA 5 CTRCTARGATCAC 4 CCTRCTARGATCA 5 CTRCTARGATCAC 4 CCTRCTARGATCAC 5 CTRCTARGCACA 4 CCTRCTARGCACA 5 CTRCTARGCACA 4 CCTRCTARGCACA 5 CTRCTARGCACA 4 CCTRCTARGCACA 5 CTRCTARGCACA 5 CTRCTARGCACAA 5 CTRCTARGCACAA 5 CTRCTARGCACAA 5 CTRCTARGCACAA 5 CTRCTARGCACAA 5 CTRCTARGCACAA 5 CTRCTARGCACAA 5 CTRCTARGCACAA 5 CTRCTARGCACAA 5 CTRCTARGCACAAA 5 CTRCTARGCACAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	Internet in the second	GGRETHIRCT REGETTERMINE REGETTERMINE RETETURET RETETURET RETETURET 2430 CREATE		1661811CACC ARACITETIC GARITAANG TIGARACCA 2450 2450 2450 2450 2450 2450 2450 2450 2450 2450 2450 2450 2450 2450 2450 2500 1000 1		IGHETIGCC11 GETTATECCH ICCTITICTGTI TICCGTIGTI TICCGTIGTI TICCGTIGTI TICCGTIGTI TICCGTIGTI TICCGTIGTI TICCGTIGTI TICCGTIGTI TIGCTIGGTI TIGCTIGGTI TIGCTIGGTI TIGCTIGGTI TICCGTIGTI TITICCTIGGGGGGG TITICCTIGGGGGG TITICCTIGGGGGG TITICCTIGGG TITICCTIGGGGGG TITICCTIGGGGGG TITICCTIGGGGGG TITICCTIGGGGGG TITICCTIGGGGGG TITICCTIGGGGGG TITICCTIGGGGGG TITICCTIGGGGGG TITICCTIGGGGGG TITICCTIGGGGGG TITICCTIGGGGGG TITICCTIGGGGGG TITICCTIGGGGGGG TITICCTIGGGGGGG TITICCTIGGGGGGG TITICCTIGGGGGGG TITICCTIGGGGGG TITICCTIGGGGGGG TITICCTIGGGGGGG TITICCTIGGGGGGG TITICCTIGGGGGGG TITICCTIGGGGGGG TITICCTIGGGGGGG TITICCTIGGGGGGG TITICCTIGGGGGGG TITICCTIGGGGGGG TITICCTIGGGGGGG TITICCTIGGGGGG TITICCTIGGGGGG TITICCTIGGGGGGG TITICCTIGGGGGGG TITICCTIGGGGGG TITICCTIGGGGGG TITICCTIGGGGGGG TITICCTIGGGGGG TITICCTIGGGGGGG TITICCTIGGGGGGG TITICCTIGGGGGGG TITICCTIGGGGGGG TITICCTIGGGGGGG TITICCTIGGGGGGG TITICCTIGGGGGGG TITICCTIGGGGGGG TITICCTIGGGGGGGGGGGG TICCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	IN	CICARAGET CARATECT IN ARAGEAGET INCARRECT INCARRECT INCARRECT INCARRECT 2490 OTGAGEGAGET OTGAGEGAGE OTGAGEGAGE OTGAGEGAGE OTGAGEGAGE OTGAGEGAGE OTGAGEGAGE OTGAGEGAGE OTGAGEGAGE OTGAGEGAGE OTGAGEGAGE OTGAGEGAGE OTGAGEGAGE OTGAGEGAGE OTGAGEGAGE OTGAGEGAGE OTGAGEGAGE OTGAGEGAGE INCOM I		GTGCTARGG GGGATTGGG CGGATTGGG CGGATTGGG GGGATGGG CTGATAGG GGGATCGGT SGGATCGGT SGGATCGGT SGGATCGGT SGGATCGGATC SGGATCGA	CHARGTGRAN THRGETGET INFGGTGET (CCAG 2520 	GGECTTEGGE CAGETACTINGGE CAGETACTINGGE CAGETACTINGGE TIGACCGGETT TIGACCGGETT SAGTATACTINGGE SAGTATACTINGGE ANTCACCCET TIGACCGGET TIGACCGGET SAGTATACTINGE ANTCACCCEG TIGATACTINGE ARTACACCEG ANTCACCCEGT TIGATACTINGE ARTACTINGE ARTACTINGE </th <th>TTTOCTARGG ANDICAN TAC TARGET CONTINUES AND</th> <th>STIRI-C TTCRA-T TTCRA-T TTCRA-T TTCRA-T TTCRA-T STGRAGC SGRCTGC SGRCTGC SGRCTGC SGRCTCC SGRCTCC SGRCTCC SGRCTCC CTGGGGC ACTCARA ATTACTA ATTACTA ATTACTA ATTACTA GRTA-CT GRTA-CT GRTA-CT GRTA-CT GRTA-CT GRTA-CT GRTA-CT GRTA-CT GRTA-CT CTGGGGC CCGGGCG CCGGGC CCGGC CCGGGC CCGGC CCGGGC CCGGC CCGGC CCGGGC CCGC CCGCGC CCGGC CCGGC CCGC CCGCGC CCGCGC CCGCGC CCGC CCGCGC C</th>	TTTOCTARGG ANDICAN TAC TARGET CONTINUES AND	STIRI-C TTCRA-T TTCRA-T TTCRA-T TTCRA-T TTCRA-T STGRAGC SGRCTGC SGRCTGC SGRCTGC SGRCTCC SGRCTCC SGRCTCC SGRCTCC CTGGGGC ACTCARA ATTACTA ATTACTA ATTACTA ATTACTA GRTA-CT GRTA-CT GRTA-CT GRTA-CT GRTA-CT GRTA-CT GRTA-CT GRTA-CT GRTA-CT CTGGGGC CCGGGCG CCGGGC CCGGC CCGGGC CCGGC CCGGGC CCGGC CCGGC CCGGGC CCGC CCGCGC CCGGC CCGGC CCGC CCGCGC CCGCGC CCGCGC CCGC CCGCGC C
HSH Consensus HSS HSS HSS HSS Consensus HSS HSS Consensus HSS HSS HSS HSS HSS HSS HSS HSS HSS HS	2 CCANARASICASIA 4 CAGG-BARMACT 5 ANGARANTACINA 2401 2410 1 CTCIGARITAC 1 CTCIGARITAC 4 CAGG-BARMACT 1 CTCIGARITAC 1 CTCIGARITAC 1 CTCIGARITAC 2401 2410 1 CTCIGARITAC 1 CTCIGARITAC 2401 2410 1 CTCIGARITAC 2551 2550 1 CTCICARASICA 2551 2550 1 CTCICARASICA		General Information Informatio Informatio Information Information Information Information	ABTGING-C GETGINGCD GETGINGCD GETGINGCD ACHIGITS CELLING 2440 AND AND AND AND AND AND AND AND AND AND	166181760C 8080176176 608117600 166081760 166081760 166081760 176081 17		Generation of the second secon	III	CCCATTCATA INCARAGE CIANAL CANANA		GTGCTARAGG GAGATTGGG CCGATTGGG CCGATTGGG CCGATTGGG GAGATCGGT ATGTGAGGAT CCGATAGGA SGATCGCAT SGATCGGATTG SGCTGGATG SGCGATG SGCATG S	CHARGTGAN THRGGTGCT INFGGTGCT INFGGTGCT ICCAG 2520 2520 2520 2520 2520 2520 2520 252	GGETCTTGGAR CAGIRCTARGE CAGIRCTARGE TGGACGGETT TGGACGGETT ABTCARCARA 2530 ABTCARCARA ABTCARCARA GENTARCTT TGTCARGET GENTARCTT TGTCARGET GGETCARGET GGETCARGET ABTCARCHART SCARACCHART CARACCART		3TTR1-C TTCRAP-T TTCRAP-T TTCRAP-T TTCRAP-T TTCRAP-T TTCRAPC SCORTCA S
HSH Consensu HSS HSS HSS HSS Consensu HSS HSS Consensu HSS HSS HSS HSS HSS HSS HSS HSS HSS HS	2 CCRAMBERITANGE 2 CCRAMBERITACIONE 6 CARGE-GRAMEATT 5 ARGARITACIONE 1	Internet in the second	GGRETHIRCT GGRETGRET REGETTGRET REGETTGRET REGETTGRET REGETTGRET REGETTGRET REGETTGRET 2430 COMMENTION 2430 COMMENTION 2430 COMMENTION 2430 COMMENTION 2580 C		166181CAC 6081TAAAG 6081TAAAG 16880CAC 16880CAC 16880CAC 16880CAC 16880CAC 16880CAC 169		GBGETGGC11 GETTATECCR GETTATECCR GETTATECCR GETTATECCR GETTATECCR GETGGGGGAGNAR 2470 GETGGGGGAGNAR GETGGGGGAGNAR GETGGGGGAGNAR GETGGGGGGAGNAR GETGGGGGGAGNAR GETGGGGGGGAGNAR GETGGGGGGGAGNAR GETGGGGGGGGAGNAR GETGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	IN	CICARAGE CARATECT IN ARAGEAGE TI INCARRECT INCARRECT INCARRECT INCARRECT INCARRECT 2490 OTGAGICARA AND AND AND AND AND AND AND AND AND AND AND AND AND AND		GTGCTARGG GGGATTGGG CGGATTGGG CGGATTGGG CGGATTGGG GGGATCGGT SCALL	CHARGTGRAN THRGETGET INGGETGET (CCGG RE ICAGGTGET (CCGG RE ICAGGTGET TCCGG TTCC TCCGG TTCCCG TTCCCG TTC	GGETCTTGGAR CAGTRETTAGGE CAGTRETTAGGE TGGACGGETT TGGACGGETT TGGACGGETT TGGACGGETT SGATARCTTG SGATARCTTG TGGACGGETT TGGACGGETT TGGACGGETT TGGACGGETT TGGACGGETT TGGACGGETT TGGACGGETT TGGACGGETT TGGACGGETT GGGACTAGET GGGACTAGET GGGGACTAGET GGGGACTAGET GGGGACTAGET GGGGGACTAGET GGGGGACGGET GGGGGACGGEGGEGGGGGGGGGGGGGGGGGGGGGGGGGG	TTTACTARAGE THACTARAGE ARRIGGTARGAN CARGATARGAN TTCASTARGAN CARGATARGAN CARGATARGAN CARGATARAGAN CARGAN CARGATARAGAN CARGA	STTRI-C TTCRA-T TTCRA-T TTCRA-T TTCRA-T TTCRA-T STGRAGC SGRCTGC SGRCTGC SGRCTGC SGRCTCC 2550

bioRxiv preprint doi: https://doi.org/10.1101/142612; this version posted May 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

	001 3010	3020	3030	3040	3050	3060	3070	3080	3090	3100	3110	3120	3130	3140	315
		GCTGGAAGTGT												TRTRRAG	
MSH7 A	ATACTTGATGAC IATACAA-ATATC	CGCTATO	CTCGAACTIT	ACTATTGACCO	GACCRAATA	TGGGTGGAAA	TCGA	-CICI-TIT	A-CGIGCCTC	TGTCTGGCT	GTCATAATGGC	TCAGTTGGG	TIGCTATGIG	CCTGGAGA-A	ACATG
HSH6 T HSH2 T	TTTGTGAACATO TTGGTTTF TCGCRATO	GCTTGGGCTAT	TTGTGRGCRC	AUTOTACTA	ACTGCHGHHT AAATTAAAGC	ACCARCATTG	INTCHGTT T-TGCAR	-CTCRCTTT	CHHGIGHITH	THE CATEGORIE	CCHHCHTGTCG CCARTGRGAR	TGGARACAR	CHHGTCHGTG TGGACATAAG	CAHGICHHG CAAATTTCCA	HTGH- GTGTG
HSH4 C HSH5 R	TCGCRATO		TTGTGAGCAT	CTATTGGC	ACTTRAAGC	GTACTCAATA	T-TGCTR	-CTCATATO	GRANATCTAT	AGCGCTA		TECRARATET			TTGA-
onsensus .	ttggtaTo	gct.ggag	aa.ca.	.tt.t.g	a.ttaaa.c	taa.t.	t.tgC.A	.CLCa.LTE	.atat	t6t	.ccagta	Tccaaft.	.aat.t.	catt.a.g	tgtg
3	151 3160	3170	3180	3190	3200	3210	3220	3230	3240	3250	3260	3270	3280	3290	33
HSH1 A		GACGGTCAACO	RATACCAACT	TGGAAACTO	ATTGATGGG	A-TCTGTAAA	GAGAGTCTAC	CATTIGAA-	ACAGCTCAGA	AGAA	GAATTCCAGA	AAT	-ATTAATCCA	AGAGCAGAA	GAATT
HSH7 T	GCTTTC8CTT61	GGACATCATC1	TTACACGTCT	T66866CC8CT0	GATCGRATCA	BATAAT 367	BAGTACCT	TCTTCATIC	BAATACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	BUTBUGUE	TTCTTCBBBB	TACATACATA	TAATTCCCI	ITGITCITTI	GATE
HSH6 I HSH2 A	GCCAGTTCTTC1 AATTTTCATGT(AGT <mark>GCACA</mark> CA1	TGACTCTTCT	AGTCGCARGCI	THERHECC AACTATECT	TTACAAGGTT	CAACCAGGTG	CTTGTGATC	ARAGTTTT		GTATTCACGT	AGCAG	-GCTTTATCCI -AGTTTGCCA	ATTITCCACA	AAGTO
MSH4 T MSH5 T	GAGGARTARCCU ACTGGAG	CATGGACTICA	AGTTTCAACT	GAAGGATGGG	CACGTCATG	TGCCACATTA	GGCCTCATG	CTAGCAGGA	IGTRGCT		GATTACCAAG	TTCA	-GTAGTGGAG	ACTECCARAR	GAATO
	getg.atat	g.aca	ca.ct	•••••	gag	tca	tg	c.t	agct	•••••••	gttc.a	c	t.tc.;	ac.a.a	ga.t.
1	301 3310	3320	3330	3340	3350	3360	3370	3380	3390	3400	3410	3420	3430	3440	34
	TAATTCAGCTTE														
HSH7 -	TTGGGTAGAGG	ACAAGCACTTI	TGATGGATAT	GCTATTGCATE	TGCTGTATT	CCGACACCTT	STAGA <mark>GAC</mark> GG	-TGARCTGC	CGCTTACTGT	TTGCCAC-ACI	ATTAICATCCA	CTCACTAAA	GAGTITGCTT	CTCATCCACA	TGTGG
MSH2 G	CATGGGAGGAAF	GGCCCTGGC	CAGAGAAAAAG	GCTTCTGAGTI	IGGAGGATIT	CICICCICGI	GCTATGA-TG	CCARATGAC	TGTARAGAGG	RAGICICAR		ACCGGAAGA	GGGRATI	ITGACCCACA	IG
MSH5	TCT														
	t	an anna canada	10 - 10 - 10 - 10 - 10 - 10 - 10 - 10 -		0.7.2.2.2.2.2.2.2.2.2.2.2.2.2.2.2.2.2.2.		NAMES OF COMMON		and the second second			N	States and the second		
ĭ	451 3460	3470	3480	3490	3500	3510	3520	3530	3540	3550	3560	3570	3580	3590	36
HSH1 T HSH3 R	TRATECA-TCG	ATGGGAATTTC	AAGCAAGAAA	-CTTGAAGATO	CTATCTOTO	TTATCTGTGA	GAAGAAGTTA	ATTGRG	CTGTATARA	GARARATCC	STCAGAAATGC	CAATGGTGA	ATTGCGTTCT	TATIGCTOCC	RGGG
MSH7 T	TACAACACATGO	CCTGCTCTTTC	AAGT-TGARA	TCTCRAAGTTO	ATCTCCAAC	AGAGCAAGAA	CTGGTGTTC-	CTCTACCGC	CTAACCTCTG	-GAGCATGCC	CGGAGAGCTAT	GGGATGCAA	GTAGCCCTTA	TGGCTGGAAT	ACCG
HSH2 T	TATGGCAGGCCC	100000000000000000	TCGTCBOTTC	TTACAGGATTI	TTAR <mark>TTART</mark> IN	GCCACT6GAT	- TR 22TR286		CT8886	aggegttgettge	ABC <mark>AGT</mark> TGAGC	CARATGARG	астолеатле	odtr agara ar	RGTTI
HSH4 T HSH5	GAAGATTCTTT	AGGGAGGC	TCTGCAGAAC	CTTAAGGAAAA	ICTATATTGG	TGGAAGGCTT	REGARARTG	TARAAG	-AGTGTAAAA	AGATGAGACI	AT- <mark>GT</mark> TGTAT	TGTTTATAC	ACAGAGTITG	TGARG-TGAG	AGAGI
	.a	ag	t.g.cagaac	t.ag.a	.ctc	c.g	aggt.		taaa.	c	gt	ta.	agatt	.g.ag.t	ag.g.
	3601 3610	3620	3630	3640	3650	3660	3670	3680	3690	3700	3710	3720	3730	3740	3
MSH3 MSH7 MSH6	A-GCCGGCTCC ATGGTAATGAC	GGAGACTCCA	TTTCTRAT	GCAAGCTAGA GTAATGAAGA	AGCCTTGCAG ARATGAACAG	CTCRGTTRAT	BAGTABA	TECTEGTAC	TGARACTGCC	GC-CIGCIT	STACAGTAGGA	CACCACTAT	TTTGATCCAR	AAGGAGGG TATAGATGT	IGAAT
HSH4	HCTGFGGTTGH TTGAACACTTTI A-GTCAGTGGC AGATAAATAAT	TCCACAAGGT	CAATGTCGAG												ATGGT
HSH4 HSH5	TTGAACACTTTI R-GTCAGTGGC	TCCACAAGGT TCAGCA TCCTARATC	CARTGTCGAG STTCTTTAGTT	CTTCAARTTR	6	CATCGGTTAT	CTATTGACT-	ATCAAAAA							ATGGT
NSH4 NSH5 Consensus	TTGAACACTTT A-GTCAGTGGC AGATAAATAAT	TCCACAAGGT TCAGCA TCCTARATC	CARTGTCGAG STTCTTTAGTT	CTTCAARTTR	6	CATCGGTTAT	CTATTGACT-	ATCAAAAA							ATGGT
HSH4 HSH5 Consensus HSH1	TTGAACACTTTA A-GTCAGTGGC AGATAAATAATA at 3751 3760 I	TCCRCRAGGT(TCRGCR(TCCTARATC ,tca) 3770 TATATTTC	SCAATGTCGAG GTTCTTTAGTT 3 3780 TTAGTCTC	3790	5 3800 CATCGCCTGC	3810 3810	3820 3820	ATCAAAAAA 3830 'AAATCA	ACTCCAGGGT 3840	3850	-TGCCATATGG 3860 CAGCTAACAAA	3870 CGTTGCTGA	ITGGAAAAGG1 3880 Iggtaagcato	TTCCGGAGGT 3890 CGTAATTTTG	ATGGT CTTGA 3 GCA
HSH4 HSH5 Consensus HSH1 HSH3	TTGAACACTTTA A-GTCAGTGAGC AGATARATAATA at 3751 3760 I	TCCACAAGGT TCAGCA TCCTAAATC .tca 3770 TATATTTC CAATTCAGA	SCAATGTCGAG STTCTTTAGTT 3780 TTAGTCTC SCTTTGGTCTA	CTTCRARTTR 3790 - TGGCRRGRG GTGGTRRGRG	G 3800 CATCGCCTGC CACAACCCGT	3810 CAATTGGARA	CTATTGACT- 3820 CTCTTCTAAT	ATCAARAA 3830 AAATCA ATCATGGT-	3840 ACTTCCTAR	3850 	-TGCCATATGG 3860 CAGCTAACAAA SACAAA-AGCA	3870 CATGCCAAA 3870 CGTTGCTGA TGATATTAA	TTGGAAAAGG1 3880 TGGTAAGCATO RCTTARGAGTA	TTCCGGAGGT 3890 CGTAATTTTG ARGGGGTGG	ATGGT CTTGA 3 GCA GCTTA
HSH4 HSH5 Consensus HSH1 HSH3 HSH7 HSH6	TTGAACACTTTA A-GTCAGTGGC AGATAAATAATA at 3751 3760 I	TCCACAAGGT TCAGCA	SCAATGTCGAG STTCTTTAGTT 3	3790 -TGGCAAGAGG GTGGTAAGAGG	3800 CATCGCCTGC CACAACCCGT CATTGTTT	3810 CCAATTGGAAAA GATTTGTGGGG TTGTTTGTGGGC	CTATTGACT- 3820 CTCTTCTAAT ITGGGCATAC	8TCAAAAAA 3830 AAATCA 8TCATGGT- 166C-TGAT-	3840 	3850 CATGGTTTTT GATGGTTTTA	-TGCCATATGG 3860 CAGCTAACAAA SACAAA-AGCA STTGGT-CTTT	3870 CGTTGCTGAT TGATATAGAC	TTGGAAAAGG1 3880 TGGTAAGCATC ACTTAAGAGTA CATTATGCTG1	3890 3890 CGTARTITIG TRAGCGCTGG	ATGGT CTTGA 3 GCA GCTTA ATGTT
HSH4 HSH5 Consensus HSH1 HSH3 HSH7 HSH6 HSH2 HSH4	TTGAACACTTTC A-GTCAGTGGC AGATAAATAAT at. 3751 3760 I	TCCACAAGGT TCAGCA	SCAATGTCGAG STTCTTTAGTT 3	3790 -TGGCAAGAGG GTGGTAAGAGG	3800 CATCGCCTGC CACAACCCGT CATTGTTT	3810 CCAATTGGAAAA GATTTGTGGGG TTGTTTGTGGGC	CTATTGACT- 3820 CTCTTCTAAT ITGGGCATAC	8TCAAAAAA 3830 AAATCA 8TCATGGT- 166C-TGAT-	3840 	3850 CATGGTTTTT GATGGTTTTA	-TGCCATATGG 3860 CAGCTAACAAA SACAAA-AGCA STTGGT-CTTT	3870 CGTTGCTGAT TGATATAGAC	TTGGAAAAGG1 3880 TGGTAAGCATC ACTTAAGAGTA CATTATGCTG1	3890 3890 CGTARTITIG TRAGCGCTGG	ATGGT CTTGA 3 GCA GCTTA ATGTT
HSH4 HSH5 Consensus HSH1 HSH3 HSH7 HSH6 HSH2 HSH4 HSH4 HSH5	TTGAACACTTTC A-GTCAGTGGC AGATAAATAAT at. 3751 3760 I	TCCACAAGGT TCAGCA	SCAATGTCGAG STTCTTTAGTT 3	3790 -TGGCAAGAGG GTGGTAAGAGG	3800 CATCGCCTGC CACAACCCGT CATTGTTT	3810 CCAATTGGAAAA GATTTGTGGGG TTGTTTGTGGGC	CTATTGACT- 3820 CTCTTCTAAT ITGGGCATAC	ATCAAAAAA 3830 AAATCA ATCATGGT- IGGC-TGAT-	3840 	3850 CATGGTTTTT GATGGTTTTA	-TGCCATATGG 3860 CAGCTAACAAA SACAAA-AGCA STTGGT-CTTT	3870 CGTTGCTGAT TGATATAGAC	TTGGAAAAGG1 3880 TGGTAAGCATC ACTTAAGAGTA CATTATGCTG1	3890 3890 CGTARTITIG TRAGCGCTGG	ATGGT CTTGA 3 GCA GCTTA ATGTT
HSH4 HSH5 Consensus HSH1 HSH3 HSH7 HSH6 HSH2 HSH4 HSH5 Consensus	TTGAACACTTTC A-GTCAGTGGC AGATAAATAAT at. 3751 3760 I	TCCACAAGGT TCAGCA	SCAATGTCGAG STTCTTTAGTT 3	3790 -TGGCAAGAGG GTGGTAAGAGG	3800 CATCGCCTGC CACAACCCGT CATTGTTT	3810 CCAATTGGAAAA GATTTGTGGGG TTGTTTGTGGGC	CTATTGACT- 3820 CTCTTCTAAT ITGGGCATAC	ATCAAAAAA 3830 AAATCA ATCATGGT- IGGC-TGAT-	3840 	3850 CATGGTTTTT GATGGTTTTA	-TGCCATATGG 3860 CAGCTAACAAA SACAAA-AGCA STTGGT-CTTT	3870 CGTTGCTGAT TGATATAGAC	TTGGAAAAGG1 3880 TGGTAAGCATC ACTTAAGAGTA CATTATGCTG1	3890 3890 CGTARTITIG TRAGCGCTGG	ATGGT CTTGA 3 GCA GCTTA ATGTT GCCAC
HSH4 HSH5 Consensus HSH1 HSH3 HSH3 HSH4 HSH4 HSH5 Consensus HSH1	TTGARGCCTTT A-GTCARCTGCC RGATARATART 3t 3751 3760 I	TCCRCRAGGT TCGGCA(TCGGCA(TCCTARATC .tca) 3770 TATATTC RATTCAGA(RATTCAGG CTTTACAGG 3920	200701000000000000000000000000000000000	3790 -TGCCARGAG GTGGTARGAG GTGGTARGAG ITTGATA- CAGGTGCATG 3940	3800 2870600000 28706000060 2870600000 2870600000000000000000000000000000000000	3810 3810 CRATTGGAAA GATTTGGGG TGTTGGGG TGTTGGGG TATG-GTGTC TATG-GTGTC 3960	3820 3820 CTCTTCTAAT TIGGGCATAC ATGAACTTAA AATGTTGCAC 3970	3830 3830 786756 7876756 8667567 8667567 8667567 3980	3840 	3850 2850 2850 2815 2815 2815 2815 2815 2815 2815 2815	-TGCCATATGG 3860 	3870 3870 CGTTGCTGAT TGATATAGACC TGCAAAAATCT TGCAAAAATCT 4020	3880 3880 IGGTANGCATI ACTTANGCATI CATTATGCTGI IGAGGCGCTTC 4030	3890 3890 2017 2017 3890 2017 2017 2017 2017 2017 2017 2017 201	ATGGT CTTGA 3 GCA GCTTA Atgtt GCCAC 4
HSH4 HSH5 Consensus HSH1 HSH3 HSH7 HSH6 HSH5 Consensus HSH1 HSH3 HSH7 HSH6 HSH2 HSH4 HSH3 HSH7 HSH6 HSH2 HSH4 HSH4 HSH4 HSH4 HSH4 HSH4 HSH4	TTGARACACTTT A-GTCRACTGAC AGATARATRATI at 3751 3760 	TCCCRARGGT TCRGCR TCCCTRRATC TCCTRRATC .tca 3770 TRTATTTC RRTTCRGA TTTACAGG 3920 GGTTTCTCC IGRARAGGGTGT	2000 TETECORE STTETTTAGTT 3780 	CTTCRAATTAU 3790 -TGCCAAGAGA ITTGATA- CAGGTGCATGA 3940 	3 3800 CATCGCCTGC CACAACCCTG CACAACCCTG CACAACCCTG CACAACCCTG 3950 3950 SAGTCAGCTT	3810 3810 CAATTGGAAA GATTTGIGGG TATG-GIGIC 3960 GGAGCATTGG GGAGCATTGG	3820 2011	3830 3830 984704 9660-7681- 9660-7681- 9660-7681- 980 3980 3980 980 980 980 980 980 980 980 980 980	3840 -ACTTCCTAA -ACTTCCTAA -TCAARCTCT -AGATCCACT GACTTCCTGA 3390 	3850 3850 10216611111 62168881111 16216888111 16316160111 4000	-TGCCATATGG 3860 - AGCTAACAAA - AGCTAACAAAA - AGCAAAA-AGCA - AGCAAGAGCTGC - 4010 - 4010 - 4010	3870 	3890 GGTARGGAT CTTARGGAT CTTARGGT TGAGGCGCCTTC 4030 TARARTCTGCT	3890 3890 2014 2014 2014 2014 2014 2014 2014 201	ATGGT CTTGA 3 GCA GCCTTR ATGTT GCCAC 4 TTTGG ATCAG
1514 1515 Concensus 1511 1513 1517 1516 1514 1515 Concensus 1511 1513 1514 1515 Concensus 1511 1514 1516 1514 1516 1514 1516 1514 1516 1514 1516	TTGARCACTTT A-GTCARETGEC RGATARATART 3t 3751 3760 I	TCCCRARGGT TCRGCR TCCCTRRATC TCCTRRATC .tca 3770 TRTATTTC RRTTCRGA TTTACAGG 3920 GGTTTCTCC IGRARAGGGTGT	2000 TETECORE STTETTTAGTT 3780 	CTTCRAATTAU 3790 -TGCCAAGAGA ITTGATA- CAGGTGCATGA 3940 	3 3800 CATCGCCTGC CACAACCCTG CACAACCCTG CACAACCCTG CACAACCCTG 3950 3950 SAGTCAGCTT	3810 3810 CAATTGGAAA GATTTGIGGG TATG-GIGIC 3960 GGAGCATTGG GGAGCATTGG	3820 2011	3830 3830 984704 9660-7681- 9660-7681- 9660-7681- 980 3980 3980 980 980 980 980 980 980 980 980 980	3840 -ACTTCCTAA -ACTTCCTAA -TCAARCTCT -AGATCCACT GACTTCCTGA 3390 	3850 3850 10216611111 62168881111 16216888111 16316160111 4000	-TGCCATATGG 3860 - AGCTAACAAA - AGCTAACAAAA - AGCAAAA-AGCA - AGCAAGAGCTGC - 4010 - 4010 - 4010	3870 	3890 GGTARGGAT CTTARGGAT CTTARGGT TGAGGCGCCTTC 4030 TARARTCTGCT	3890 3890 2014 2014 2014 2014 2014 2014 2014 201	ATGGT CTTGA 3 GCA GCCTTR ATGTT GCCAC 4 TTTGG ATCAG
HSH4 HSH5 Concensue HSH1 HSH3 HSH7 HSH6 HSH5 Concensue HSH1 HSH3 HSH7 HSH6 HSH3 HSH7 HSH6 HSH6 HSH6 HSH6 HSH6 HSH6 HSH6	TTGARCACTTT A-GTCARETGEC RGATARATART 3t 3751 3760 I	TCCCRARGGT TCRGCR TCCCTRRATC TCCTRRATC .tca 3770 TRTATTTC RRTTCRGA TTTACAGG 3920 GGTTTCTCC IGRARAGGGTGT	2000 TETECORE STTETTTAGTT 3780 	2790 -TGCARGROUT GTGCTARGROUT GTGCTARGROUT GTGCTARGROUT CRGGTGCCATG 3940 	3 3800 CATCGCCTGC CACAACCCTG CACAACCCTG CACAACCCTG CACAACCCTG 3950 3950 5AGTCAGCTT	3810 3810 CAATTGGAAA GATTTGIGGG TATG-GIGIC 3960 GGAGCATTGG GGAGCATTGG	3820 2011	3830 3830 984704 9660-7681- 9660-7681- 9660-7681- 980 3980 3980 980 980 980 980 980 980 980 980 980	3840 -ACTTCCTAA -ACTTCCTAA -TCAAACTCT -AGATCCACT GACTTCCTGA 3390 	3850 3850 10216611111 62168881111 16216888111 16316160111 4000	-TGCCATATGG 3860 - AGCTAACAAA - AGCTAACAAAA - AGCAAAA-AGCA - AGCAAGAGCTGC - 4010 - 4010 - 4010	3870 	3890 GGTARGGAT CTTARGGAT CTTARGGT TGAGGCGCCTTC 4030 TARARTCTGCT	3890 3890 2614811116 1806561562 2604811116 26049 20040 4040 168020868	ATGGT CTTGF 3 GCA GCCA GCCAC ATGTT GCCAC
HSH4 HSH5 Consensue HSH1 HSH3 HSH7 HSH6 HSH5 Consensue HSH1 HSH5 HSH4 HSH5 Consensus HSH4 HSH5 Consensus	TTGARACACTTT R-GTCRGTGGC RGATAGATAGATA 3t	TICCRCINGET TICCTANABCC TICCTANABCC TICCTANABCC 3770 TITTTIC TITTTCA ANTTORGA- ANTTORGA- ANTTORGA- SALAGE 3320 SGTTCTCC SGANAGCGTCT GANAACCCTTI	SCARTETCORE STICTTRAFT 3780 	2790 -TGCARGROUT GTGCTARGROUT GTGCTARGROUT GTGCTARGROUT CRGGTGCCATG 3940 	3 3800 CATCGCCTGC CACAACCCTG CACAACCCTG CACAACCCTG CACAACCCTG 3950 3950 5AGTCAGCTT	3810 3810 CAATTGGAAA GATTTGIGGG TATG-GIGIC 3960 GGAGCATTGG GGAGCATTGG	3820 2011	3830 3830 984704 9660-7681- 9660-7681- 9660-7681- 980 3980 3980 980 980 980 980 980 980 980 980 980	3840 -ACTTCCTAA -ACTTCCTAA -TCAAACTCT -AGATCCACT GACTTCCTGA 3390 	3850 3850 10216611111 62168881111 16216888111 16316160111 4000	-TGCCATATGG 3860 - AGCTAACAAA - AGCTAACAAAA - AGCAAAA-AGCA - AGCAAGAGCTGC - 4010 - 4010 - 4010	3870 	3890 GGTARGGAT CTTARGGAT CTTARGGT TGAGGCGCCTTC 4030 TARARTCTGCT	3890 3890 2614811116 1806561562 2604811116 26049 20040 4040 168020868	ATGGT CTTGF 3 GCA GCCA GCCAC ATGTT GCCAC
HSH4 HSH5 Concensue HSH1 HSH3 HSH7 HSH6 HSH5 Consensue HSH1 HSH5 Consensus HSH1 HSH5 KSH5 KSH5 HSH5 HSH5 HSH5 HSH5 HSH5	TTGARACACTTT R-GTCRGTGGC RGATAGATAGATA 3t	TCCTCARAGET TCCTCARAGET TCCTCARAGET TCCTCARAGET 3770 TATATITCO	CRAFTCCGAG STTCTTAGTT 3780 	271CRARTTAN 3790 -TGSCARGAGI GTGSTARGAGI GTGSTARGAGI 23940 	3 3800 CATCGCCTGC CACAACCCTG CACAACCCTG CACAACCCTG CACAACCCTG 3950 3950 5AGTCAGCTT	3810 3810 CAATTGGAAA GATTTGIGGG TATG-GIGIC 3960 GGAGCATTGG GGAGCATTGG	3820 2011	3830 3830 984704 9660-7681- 9660-7681- 9660-7681- 980 3980 3980 980 980 980 980 980 980 980 980 980	3840 -ACTTCCTAA -ACTTCCTAA -TCAAACTCT -AGATCCACT GACTTCCTGA 3390 	3850 3850 10216611111 62168881111 16216888111 16316160111 4000	-TGCCATATGG 3860 - AGCTAACAAA - AGCTAACAAAA - AGCAAAA-AGCA - AGCAAGAGCTGC - 4010 - 4010 - 4010	3870 	3890 GGTARGGAT CTTARGGAT CTTARGGT TGAGGCGCCTTC 4030 TARARTCTGCT	3890 3890 2614811116 1806561562 2604811116 26049 20040 4040 168020868	ATGGI CTTGA GCA GCCA GCCA GCCA GCCA GCCA GCCA

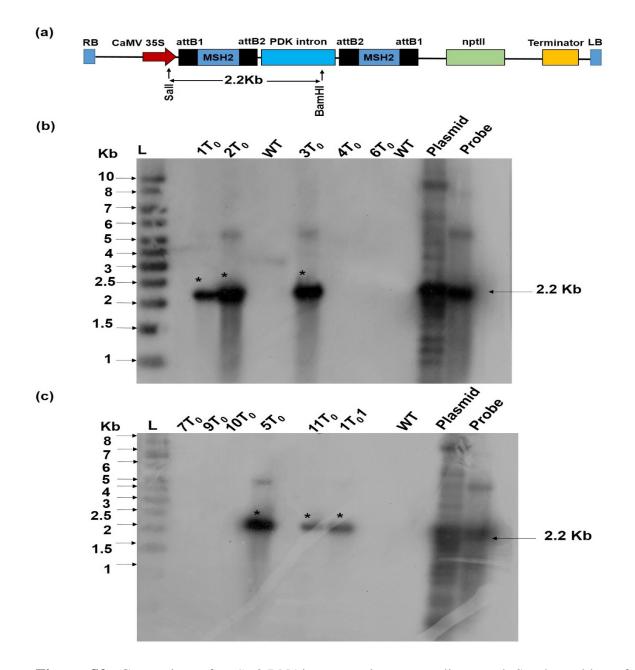


Figure S3. Generation of *MSH2*-RNAi transgenic tomato lines and Southern blot of T_0 transgenic lines. (a) Schematic representation of the construct used for *MSH2* silencing. The construct contains one spliceable intron with the targeted *MSH2* sequence forming a hairpin when the construct undergoes transformation. *AttB1* and *attB2* represent the two short stretches of sequences that participate in the recombination reaction of the Gateway system. T35S indicates 35S terminator. The restriction sites for BamHI and SalI are also indicated. (b-c) Southern blot of T_0 transgenic lines. Genomic DNA of T_0 plants was digested with BamHI and SalI to release the insert. The Southern blot was probed with radiolabelled *NPTII–NOS* probe of 2.2 Kb size. Numbers on the top of lanes $1T_0$ to $6T_0$ in panel (b) and lanes $7T_0$ to $1T_0 1$ in panel (c) indicate the plant number of respective T_0 lines. L- 1 Kb DNA Ladder (Fermentas). WT-Wild type genomic DNA (negative control); **Plasmid-**, Plasmid DNA bearing *MSH2*-RNAi construct. **Probe-** Radiolabelled NPTII–NOS 2.2 Kb probe. Note: The *NPTII–NOS* probe was obtained by digesting *MSH2*-RNAi plasmid with BamHI and SalI. The asterisk (*) indicates the presence of *NPTII–NOS* sequence in the transgenic lines.

(a)

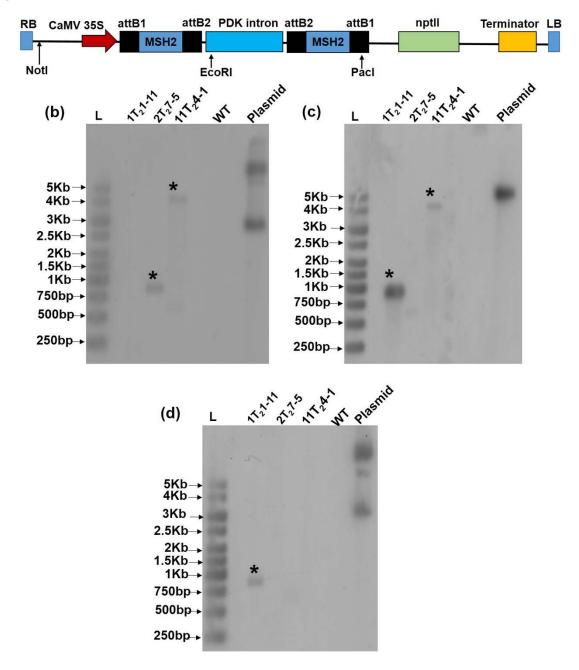


Figure S4. Southern blot of T_2 *MSH2*-RNAi lines $1T_21-11$, $2T_27-5$ and $11T_24-1$. (a) Schematic representation of the construct used for *MSH2* silencing. Black arrow represents the site of the restriction enzyme used for digesting Genomic DNA in Blot (b), (c) and (d). DNA was digested with EcoRI (b), with NotI (c), and PacI (d). L-1 Kb DNA Ladder (Fermentas). WT- Wild-type genomic DNA (negative control); **Plasmid**-, Plasmid DNA bearing *MSH2*-RNAi construct. The blots were probed with radiolabelled *NPTII* sequence. The asterisk (*) indicates the presence of *NPTII* sequence in the transgenic lines.

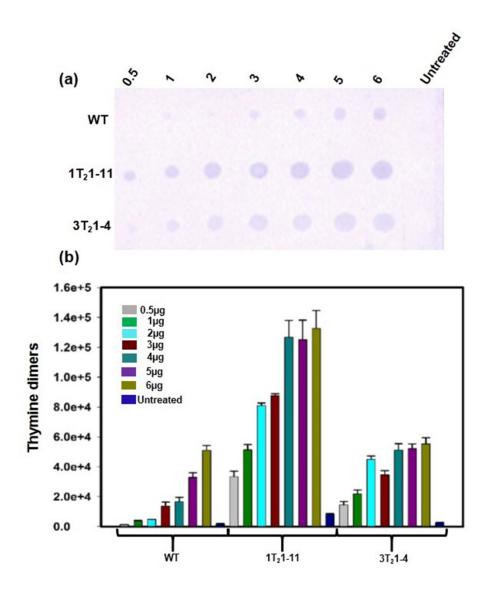


Figure S5. Quantification of thymine dimer levels in genomic DNA of WT and UV-B-treated *MSH2*-RNAi lines, $3T_21$ -4 and $1T_21$ -11. (a) Quantification of thymine dimer with different dilutions of genomic DNA. Numbers on top of the blot indicate the spotted amount of genomic DNA (µg) and on the left indicate plant number. (b) Quantification of thymine dimer formation in *MSH2*-RNAi lines in (a) by ImageJ analysis.

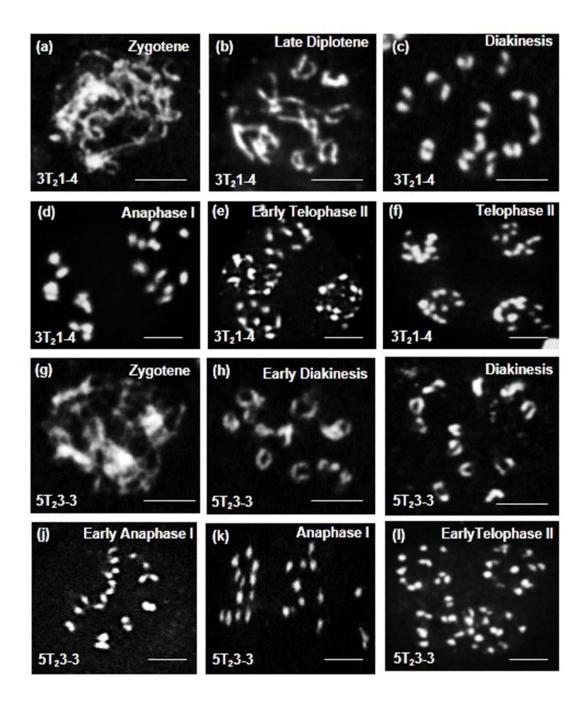


Figure S6. Male meiosis in *MSH2*-RNAi tomato lines. Representative meiotic stages of $3T_21-4$ (**a–f**) and $5T_23-3$ (**j-q**) lines from zygotene to telophase. The $3T_21-4$ line with moderate *MSH2* silencing and $5T_23-3$ line with no *MSH2* silencing displayed diploid meiotic chromosomes. Scale bar, 10 µm.

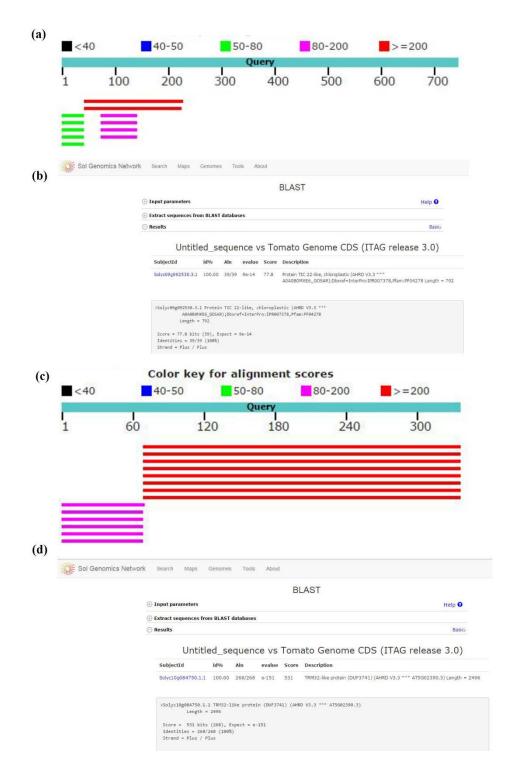


Figure S7. Sequence analysis of FPNI-PCR product of Line No. $1T_21-11$ and $2T_25-5$. The FPNI-PCR product from Line $1T_21-11$ showed 98% homology to tomato chromosome 9 in NCBI gene database (a) to *Tic22* gene of tomato in SGN database (b). The FPNI-PCR product from Line $2T_25-5$ showed 100% homology in chromosome 10 in NCBI database (c) and to *TRM32* gene of tomato in SGN database (d).

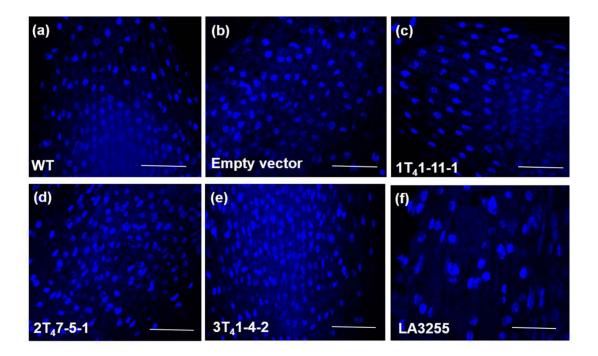


Figure S8. DAPI staining of root tip nuclei of WT and different *MSH2*-RNAi lines. (a) WT, (b) empty vector control, (c-e) progenies of $T_3 MSH2$ -RNAi lines- $1T_21$ -11-1 (c), $2T_27$ -5-1 (d), $3T_2$ -4-2 (e) and LA3255 (f). Tetraploid tomato line LA3255 shows distinct enlarged tetraploid root tip nuclei in comparison to normal diploid nuclei in WT, control and *MSH2*-RNAi lines. Scale Bar, 50 µm.

bioRxiv preprint doi: https://doi.org/10.1101/142612; this version posted May 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Table S1: The gene sequences flanking right border of T-DNA obtained using FPNI-PCR. Bold lowercase letters indicate bacterial right border T-DNA sequence and uppercase letters indicate flanking genome sequence from transgenic *MSH2*-RNAi plants.

Line No.	Right Border	FNPI-PCR sequence data
	T-DNA	
	sequence	
1T ₂ 1-11	RB	gtttattagaataatcggatatttaaaagggcgtgaaaaggtaaatgacaactatacatctgattgat
2T ₂ 5-5	RB	gtttacccgccaatatatcctgtcaaacactgatagtttaaactgaaggcgggaaacgacaatctgatcTCAGA AACACTCGATACAGCTGCTTCGACTGCATCTAGCAGCAACTTATATTCAG GGGCTACCAGGTCCAATGAGCAAAAGTCCCTGGATATTCCTTTAGGTTCA GAAAAGAAAA

Supplementary Table S2 List of genes and the primers used for screening for mutations and SNPs for the regions predicted by CODDLE based on genome sequence of Solanaceae Genome version 2.5. The *MSH2* gene information and location of the primers on genome sequence is also indicated.

Gene		Primer Sequence (5'-3')	Start	End	GC%	Tm	Amplicon
Gene		Timer Sequence (5 -5)	Position	Position	GC /0	1 111	size
M13	Fp	TGTAAAACGACGGCCAGT		1	50	53	
WI15	Rp	AGGAAACAGCTATGACCAT			42	53	
MSH2	Fp	TGTACCAATGTGCATTTTCTTCTT	42994468	42994445	33	58	
Solyc06g069230 (Set I)	Rp	TAGCTAAGAAAAGAGGGGGATTCAA	42993248	42986063	38	60	1243 bp
MSH2	Fp	CTGACACACAATCTTGAGAGGA	42993225	42993378	45	60	
Solyc06g069230 (Set II)	Rp	CAGTAGCACATCCAACTCAGAA	42992510	42992489	45	60	914 bp
MSH2	Fp	CTTCTTTCCCAGCTCATTCATC	42992585	42992563	45	60	
Solyc06g069230 (Set III)	Rp	GGGGTGGACTTTTAAGACATCA	42991635	42984452	45	60	972 bp

Supplementary Table S3: List of genes and the primers used for transcript analysis of MMR pathway.

Gene	Gene sequence id.		Primer Sequence (5'->3')	Start position	End position	GC%	Tm	Amplicon size
MSH1	Solyc09g090870.2	Fp	GAGGAACTAAAGGGGAGATTTTGT	70267311	70267288	42	62	136 bp
WISIII	(SGN)	Rp	TGAATCTAGAGCAGGTCTGAGTTG	70267199	70267175	46	63.6	150 Op
MSH2	Solyc06g069230	Fp	TTAGCTGGTTCACTTTCTGAGTTG	42992524	42992501	42	62	102 bp
1/10112	(SGN)	Rp	ATCTGGTGGACTGATATTTGGTCT	42992446	42992425	42	62	102 op
MSH3	XM_010319297	Fp	TTCATAGAAAAGCTACTGCTGCTG	1814	1837	42	62	100 bp
WISHIS	(NCBI)	Rp	ACTTCTGTCCTCCGTAATGAAAAG	1891	1914	42	62	100 Up
MSH4	XM_010326283	Fp	AGAGATGAAGGAGACGGCTTTTAT	1886	1909	42	62	139 bp
1/15114	(NCBI)	Rp	TAAGTGCCAATAGATGCTCACAAC	2002	2025	42	62	139 op
MSH5	XM_010329008	Fp	GTAGACGTCTCTTGAGGAACTGGT	1088	1111	50	65.2	119 bp
WISHIS	(NCBI)	Rp	CACGTAAAGAGACCGAAACTTCTT	1184	1207	42	62	117 op
MSH6	Solyc01g079520.2	Fp	AAGGTTTAACTGGTGGTCAGAGAC	78582874	78582893	46	65.3	108 bp
WISHO	(SGN)	Rp	GCATCCATCTCATAAAGCTCATAG	78583173	78583196	42	62	108 Up
MSH7	XM_004242880	Fp	GGATTTGCTTTTGTCGATTGTG	1499	1520	41	58.4	116 bp
141.5117	(NCBI)	Rp	TATAACTTCCTTCGGTGACACTTG	1592	1615	42	62	110 0p

Supplementary Table S4: List of genes and the primers used for FNPI-PCR. The common region is highlighted in bold.

Primer	Primer sequence (5'-3')	Primer use
FP1	GTAATACGACTCACTATAGGGCACGCGTGGT NTCGASTWTSGWGTT	1st PCR primer
FP2	GTAATACGACTCACTATAGGGCACGCGTGGT NGTCGASWGANAWGAA	1st PCR primer
FP3	GTAATACGACTCACTATAGGGCACGCGTGGT WGTGNAGWANCANAGA	1st PCR primer
FP4	GTAATACGACTCACTATAGGGCACGCGTGGT AGWGNAGWANCAWAGG	1st PCR primer
FP5	GTAATACGACTCACTATAGGGCACGCGTGGT NGTAWAASGTNTSCAA	1st PCR primer
FP6	GTAATACGACTCACTATAGGGCACGCGTGGT NGACGASWGANAWGAC	1st PCR primer
FP7	GTAATACGACTCACTATAGGGCACGCGTGGT NGACGASWGANAWGAA	1st PCR primer
FP8	GTAATACGACTCACTATAGGGCACGCGTGGT GTNCGASWCANAWGTT	1st PCR primer
FP9	GTAATACGACTCACTATAGGGCACGCGTGGT NCAGCTWSCTNTSCTT	1st PCR primer
FSP1	GTAATACGACTCACTATAGGGC	2nd PCR primer
FSP2	ACTATAGGGCACGCGTGGT	3rd PCR primer
PRB1	TTGACAGGATATATTGGCGGGT	1st PCR primer
PRB2	AAGGGCGTGAAAAGGTTTATC	2nd PCR primer
PRB3	CCATTTGTATGTGCATGCCAA	3rd PCR primer