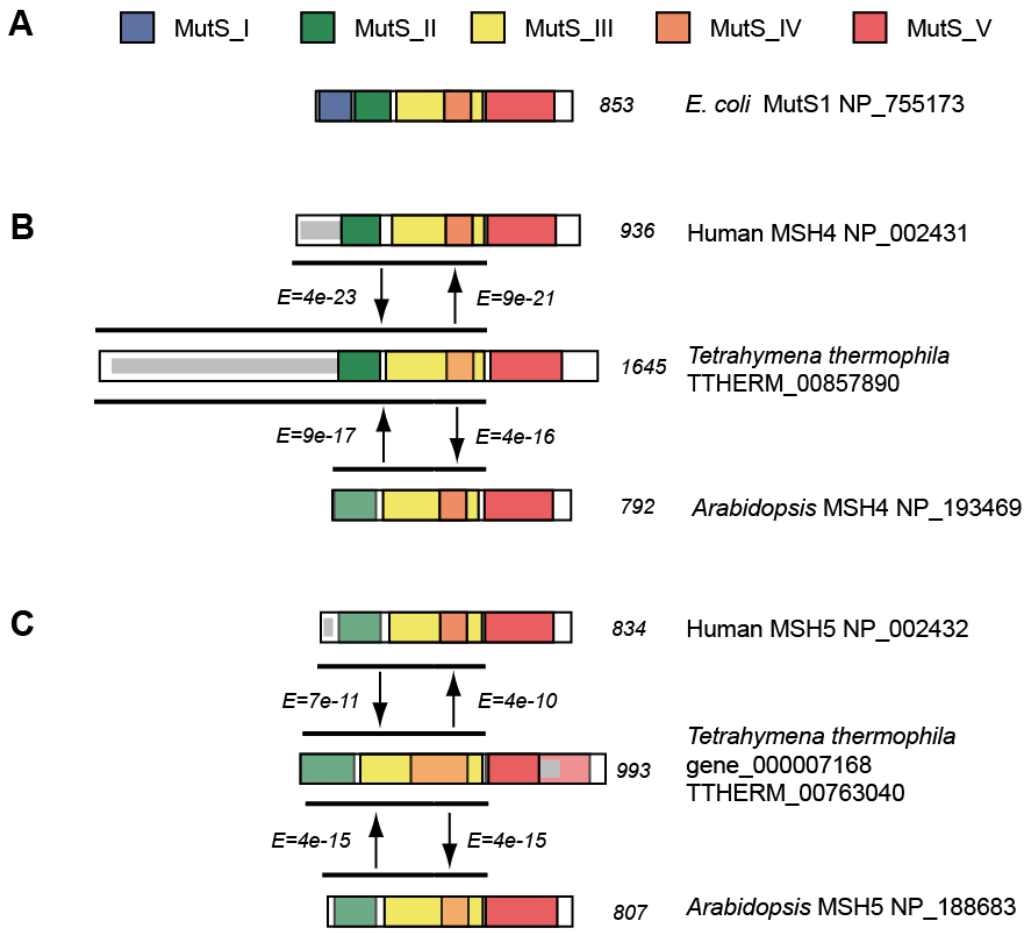

Figure S2 Comparison of MutS domain organizations. Pfam domains MutS_I to V are colored blue, green, yellow, orange, and red, respectively. Full opacity indicates a significant match in Pfam search, while reduced opacity indicates that HHpred search against Pfam was used to identify the domain. Compositionally biased regions as detected using CAST (Promponas et al. 2000, Bioinformatics 16: 915-922) are indicated in grey.

(A) *E. coli* MutS1 protein, which is composed of 5 structural domains- Muts_I-V (Obmolova et al. 2000, Nature 407: 703-710), is the likely ortholog of the eukaryotic mismatch repair (MMR) proteins MSH1, MSH2, MSH3, MSH6, as well as the non-MMR proteins MSH4 and MSH5 (Lin et al. 2007, Nucl. Acids Res. 35: 7591-7603). While MSH4 and MSH5 proteins lack a N-terminal MutS_I domain, which has been implicated in mismatch recognition, the remaining structural domains are retained with highest conservation in the MutS_V ATPase and HTH domain.

(B) Probable *Tetrahymena* MSH4 homolog, identified using Reciprocal Best Blast Searches, after excluding the C-terminal region of highest conservation among MutS homologs. Searches were performed against the human and *Arabidopsis thaliana* proteomes obtained from the NCBI non-redundant database. The *T. thermophila* proteome was derived by predicting all open reading frames longer than 300 bp from the TetraFGD RNAseq transcriptome. Regions used for BLAST are indicated by horizontal bars and E values are indicated.

(C) Probable *Tetrahymena* MSH5 homolog, identified using Reciprocal Best Blast Searches as above.