DNA mismatch repair and mutation avoidance in the ciliate protozoan *Tetrahymena thermophila*

by

Shawn Richard Salsiccioli
B.Sc., University of Victoria, 2005

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Biochemistry and Microbiology

Supervisory Committee

DNA mismatch repair and mutation avoidance in the ciliate protozoan *Tetrahymena thermophila*

by

Shawn Richard Salsiccioli
B.Sc., University of Victoria, 2005

Supervisory Committee

Dr. Claire G. Cupples, (Department of Biochemistry and Microbiology)
Supervisor

Dr. Robert D. Burke, (Department of Biochemistry and Microbiology)
Departmental Member

Dr. Caren C. Helbing, (Department of Biochemistry and Microbiology)
Departmental Member

Dr. John S. Taylor, (Department of Biology)
Outside Member

Abstract

Supervisory Committee

Dr. Claire G. Cupples, (Department of Biochemistry and Microbiology)
Supervisor

Dr. Robert D. Burke, (Department of Biochemistry and Microbiology)
Departmental Member

Dr. Caren C. Helbing, (Department of Biochemistry and Microbiology)
Departmental Member

Dr. John S. Taylor, (Department of Biology)
Outside Member

The DNA of all organisms is continuously exposed to exogenous and endogenous genotoxic agents. Fortunately, through the concerted actions of several DNA repair and mutation avoidance pathways, DNA damage can be removed and an organism's genomic stability maintained. DNA base-base mismatches are generated as a result of the inherent replication errors made by the DNA replication machinery, as well as during the meiotic pairing of homologous but non-identical chromosomes. Through the coordinated actions of the highly conserved DNA mismatch repair (MMR) system, these errors are detected, removed and corrected, thus restoring the integrity of the DNA. In the absence of DNA MMR, genetic instability is unavoidable, resulting in the accumulation of mutations, and in mammals, a susceptibility to cancer.

To better understand the roles of the MMR system in mutation avoidance during DNA replication, meiosis, and in nuclear apoptosis, we have utilized the nuclear dimorphic, ciliate protozoan *Tetrahymena thermophila*. We have identified seven putative MMR homologues; two are similar to eukaryotic *MLH1* and *PMS2*, respectively, and five are

similar to eukaryotic *MutS* homologues, one with eukaryotic *MSH2* and four with *MSH6*. Our studies demonstrate that during conjugation, the relative transcript abundance of each MMR homologue is increased compared to vegetatively growing or nutritionally deprived (starved) cells. Also, the expression profile throughout conjugation is bimodal, corresponding to micronuclear (MIC) meiosis and macronuclear (MAC) anlagen development, both periods in which DNA replication occurs. Cells containing macronuclear knockouts of the *PMS2*, *MSH2* and *MSH6_1* genes were unable to successfully pair and complete conjugation, but were viable throughout vegetative growth. Cells in which the macronuclear *MSH6_2* gene was knocked out had a phenotype that was similar to wild-type cells, during conjugation and vegetative growth. Interestingly, we observed that the MIC of cells containing MAC knockouts of the *PMS2* and *TML1* genes appear to have decreased copy number of specific "target sequences", as determined by qPCR using the Random Mutation Capture (RMC) assay. This decrease reflects neither a loss of micronuclei nor a reduction in total micronuclear DNA content.

These studies demonstrate that the *PMS2*, *TML1*, *MSH2*, and *MSH6_1* homologues are necessary for the maintenance of micronuclear function and stability during conjugal development and vegetative growth, whereas the remaining *MSH6* homologues have less pronounced roles in DNA repair and development. Additionally, macronuclear development in *Tetrahymena* appears less reliant on the DNA mismatch repair system and perhaps uses alternate surveillance mechanisms to maintain genomic stability during asexual and sexual development.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

6-mp	6-methylpurine
aa	amino acid
ATM	ataxia telangiectasia mutated
ATR	ataxia telangiectasia and Rad3-related
BER	base excision repair
BSA	bovine serum albumin
cDNA	complementary deoxyribonucleic acid
Dam	DNA adenine methyltransferase
DAPI	4',6-diamidino-2-phenylindole
DNA	deoxyribonucleic acid
DSB	double-strand breaks
EDTA	ethylenediaminetetraacetic acid
EXO	exonuclease
gDNA	genomic deoxyribonucleic acid
HNPCC	hereditary non-polyposis colorectal cancer
HR	homologous recombination
IDL	insertion/deletion loop
IES	internally eliminated sequence
kb	kilobase
KO	knockout
MAC	macronuclear, macronucleus
MIC	micronuclear, micronucleus
mg	milligram
ml	millilitre
MLH	mutL homologue
mM	millimolar
MMR	mismatch repair
MSH	mutS homologue
Mut	mutator

NER	nucleotide excision repair
NHEJ	non-homologous end-joining
NJ	neighbor-joining
nm	nanometre
PCNA	proliferating cell nuclear antigen
PCR	polymerase chain reaction
pm	paromomycin
PMS	postmeiotic segregation increased
PND	programmed nuclear death
psi	pounds per square inch
qPCR	quantitative polymerase chain reaction
rDNA	ribosomal DNA
RE	restriction endonuclease
RFC	replication factor C
RMC	random mutation capture
RNA	ribonucleic acid
rNMP	ribonucleotide monophosphate
ROS	reactive oxygen species
RPA	replication protein A
rpm	revolutions per minute
SAM	S-adenosyl methionine
SSB	single-stranded binding protein
ssDNA	single-stranded deoxyribonucleic acid
SPP	super proteose peptone
TCR	transcription coupled repair
TGD	*Tetrahymena* Genome Database
TIGR	The Institute for Genomic Research
TML1	*Tetrahymena* MutL homologue 1
U	unit
UV	ultraviolet
WT	wild-type

Acknowledgments

This thesis is the culmination of several years of work that would not have been possible without the continual support and guidance of many people, to whom I am indebted. This opportunity has provided me with invaluable practical skills and theoretical knowledge that I know will be of benefit to me in the years to come. I would first like to thank my graduate supervisor, Dr. Claire Cupples, without whose support and guidance this project would not have been possible. I am very much appreciative of the time she has generously given to me over the years. Her helpful discussions and continual encouragement have been instrumental in the completion of this thesis. I'd also like to thank the members of my graduate supervisory committee, Dr. Robert Burke, Dr. Caren Helbing and Dr. John Taylor, for their suggestions and critical discussions at committee meetings. Their expertise has been invaluable in the completion of this thesis.

I am grateful to previous colleagues in the Cupples lab whose help and fruitful discussions have unquestionably contributed to this completion of this work: Derek Bell, Lin Sun, Erin Annandale, Yaroslava Polosina and Kate Kudynska. I am indebted to Kate for her continued support and friendship. She has proven to me that regardless of whatever barrier stands in your way, anything is possible when you have the drive and passion to make it so.

I would like to give a special thanks to the graduate secretary Melinda Powell, and my graduate advisor, Dr. Steve Evans. Melinda has been a lifesaver during my

studies and made it that much easier for me to deal with any administration issues at UVic while completing my research at Simon Fraser University. She has kindly and generously gone out of her way multiple times to make sure that things go as smoothly as possible for me. Dr. Evans has been a fundamental support system to me throughout the last few years. He has always had my best interests in mind and his support and helpful discussions will always be appreciated.

Last, but not least, I would like to thank my family and friends. My parents have been a continual source of support, love and encouragement, and they never cease to show me how much they care. I can't thank them enough for always being there for me, be it if I was going through a difficult time, or if I just needed someone to throw a few ideas at, even if they didn't really understand what I was talking about. I'm also appreciative for their unending need to make sure I've been well fed; I could always expect to have a freezer full of food whenever they or anyone else came to visit. Christa Salsiccioli, aka Lupu and the best sister a brother could have, has been a tremendous source of encouragement and support. Our almost nightly phone calls will always be remembered as a welcome break. The passion she has for her work never ceases to inspire me. Finally, I would like to give deep-hearted thank you to my close friend Patricia Wallis. She and I have been through it all together. We started this journey together as undergraduates and now we finish together as graduates. Her friendship has been invaluable to me and I could not have done this without her. I thank her for making my transition from Victoria to Burnaby an easy one and for always being there to support me and listen. I wish her only the best in the years come.

Dedication

This thesis is dedicated to my parents, Anita and Rich Salsiccioli, my twin sister, Christa Salsiccioli and close friend, Patricia Wallis. Your love and support mean more than you'll ever know.

Chapter 1: Introduction

"We totally missed the possible role of … [DNA] repair… DNA is so precious that probably many distinct repair mechanisms would exist. Nowadays one could hardly discuss mutation without considering repair at the same time."

- Francis Crick, April 26, 1974 (1)

Sixty years ago, in 1953, three hallmark papers describing the structure of the DNA double helix were concurrently published in *Nature* (2-4). While controversial in their own right, these discoveries led to numerous areas of scientific investigation and discovery, perhaps the most important of which is the relationship of DNA damage and mutation to disease. In fact, as expressed in the quotation above from Francis Crick given two decades later, the implications of these discoveries were unrealized and underestimated at the time (1). We now know that in all prokaryotic and eukaryotic organisms studied to date, there are numerous dedicated repair systems in place that are able to detect, remove and correct most types of DNA damage, regardless of their source (5). One such repair system is the DNA mismatch repair (MMR) system; it is responsible for the repair of base-base mismatches and insertion/deletion loop (IDLs) that arise during DNA replication and homologous recombination (6, 7). As such, in the absence of MMR, any errors arising during these processes are left unrepaired, resulting in mutation accumulation and, in humans, an inherited susceptibility to the development of a variety of cancers, including colorectal, ovarian and uterine cancers - collectively known as Lynch syndrome (8).

To expand our understanding of MMR in eukaryotic organisms, our studies are directed at determining the contribution of DNA mismatch repair to mutation avoidance and genetic stability in the nuclear dimorphic, ciliate protozoan *Tetrahymena thermophila*. Particular attention has been focused on determining the roles of the MutS and MutL MMR homologues during the sexual phase of the life cycle, a highly complex process involving DNA replication, meiosis and apoptosis; processes in which the MMR proteins are known to act in other organisms (9).

1.1 DNA Damage and Repair

1.1.1 Endogenous and Exogenous Sources of DNA Damage

In all living cells, the maintenance of genomic integrity is essential to the continuance of life. Yet, on a daily basis, the genome is bombarded with numerous insults that can damage the chemical composition and structure of the DNA double helix. If this damage isn't repaired, the resultant mutations can lead to disease and in the worst case scenario, death. While most mutations are neutral and have no effect on an organism, some may either be beneficial or detrimental to the fitness of the organism and thus help to drive evolutionary change and variation. Damage to the DNA can arise from multiple sources (Figure 1). Exogenous agents are those originating from the external environment and include viruses, genotoxic chemicals (e.g. alkylating agents found in cigarette smoke (10)), and ultraviolet (UV) and ionizing radiation (5). Endogenous sources of DNA damage include errors made during DNA replication and meiotic recombination, aberrant methylation of the DNA bases by *S*-adenosyl methionine (SAM) (11) and oxidation of the DNA bases (i.e. C to U) by reactive oxygen species (ROS), such as those produced as

by-products of cellular metabolism (12). Additionally, deamination (13) and depurination (14) of the DNA bases can result in the creation of transition (A↔G, C↔T) or transversion (A↔C or T, G↔C or T) mutations.

1.1.2 DNA Repair and Mutation Avoidance Pathways

DNA damage resulting from exposure to exogenous and endogenous genotoxic agents necessitates repair, otherwise it will become fixed within the DNA, thus creating a mutation. To minimize the accumulation of potentially harmful mutations, prokaryotic and eukaryotic organisms have evolved complexes of proteins that act in concert to remove and subsequently repair damaged DNA bases and nucleotides (Figure 1). The simplest and most efficient of the repair mechanisms is the direct reversal of DNA damage, an example of which is the photoreactivation of cylobutane pyrimidine dimers and pyrimidine-pyrimidone (6-4) photoproducts caused by UV light. Through the activity of only a single enzyme, DNA photolyase, DNA damage can be directly reversed and the integrity of the DNA restored (15, 16). When the DNA bases are damaged by oxidation, deamination or alkylation, they are generally repaired by the base-excision repair (BER) pathway. This is accomplished through the activity of DNA glycosylases, DNA repair proteins that catalyze the hydrolysis of the N-glycosylic bond linking the damaged base to the deoxyribose-phosphate backbone (17). Upon removal of the damaged base and incision of the resultant abasic site, repair synthesis occurs followed by ligation of the DNA ends, thereby restoring the DNA to its original state.

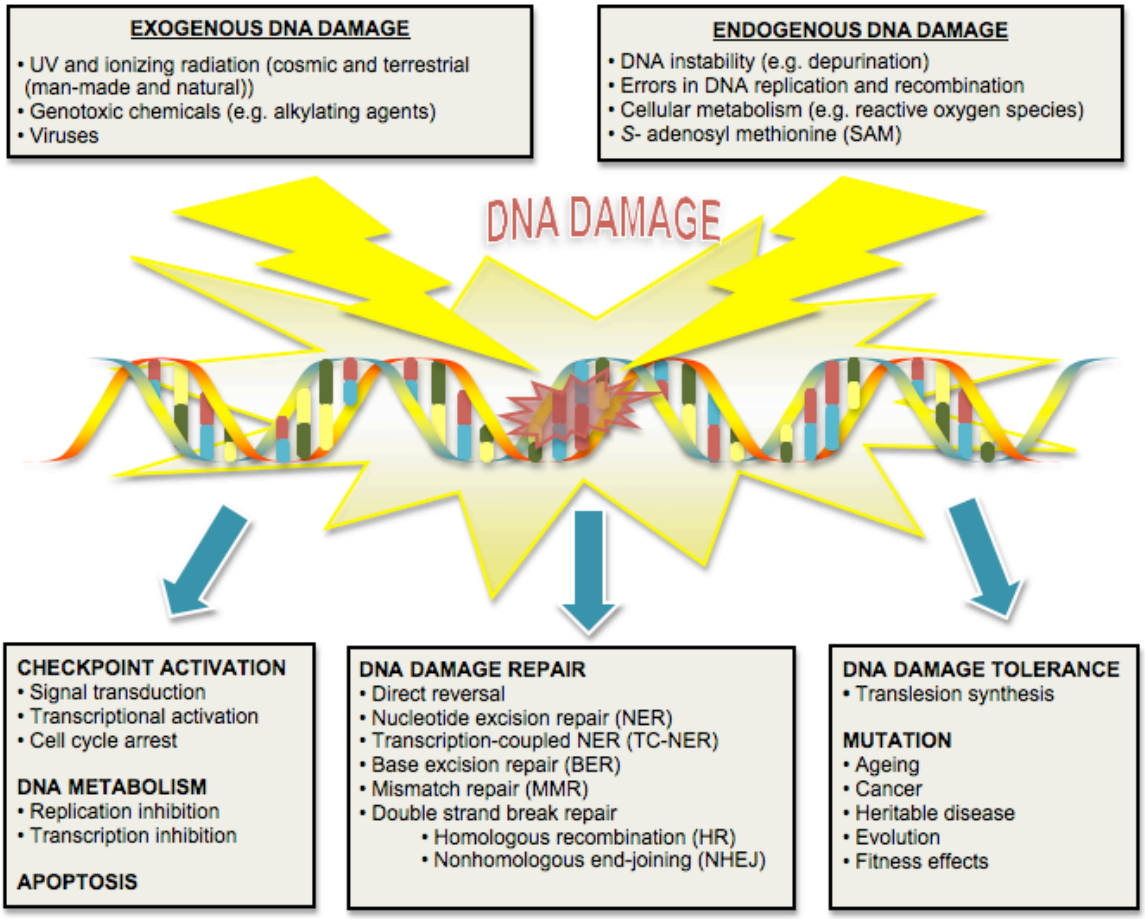


Figure 1. Sources and consequences of DNA damage.

DNA damage occurs as a result of exposure of the DNA to UV and ionizing radiation, as well as genotoxic chemicals originating in the exogenous or endogenous environments. Genomic surveillance mechanisms within the cell detect this damage and signal for the activation of various biological pathways. Most damage will be repaired by one of several DNA repair pathways and once detected may also lead to activation of the cell cycle checkpoints. If the DNA damage burden is too great for the cell to repair, or the lesion unrepairable, apoptosis will be induced, resulting in cell death. DNA damage that is not repaired, or tolerated by the cell, will be permanently fixed within the DNA as a mutation, which can have a variety outcomes including ageing and the development of cancer (adapted from (18) and rndsystems.com).

Helix-distorting bulky DNA adducts are unable to be repaired by the BER pathway are instead repaired by the nucleotide excision repair (NER) pathway. This complex repair system consists of approximately 30 proteins that together act to survey the DNA and remove damage that may block DNA replication or transcription (in this case it is called

transcription-coupled repair (TCR)) (19). TCR is different in that it acts on actively transcribed regions of the genome, with damage recognition being mediated by a stalled RNA polymerase. Though the mechanisms of damage recognition are different between these two NER pathways, both eventually converge to excise and repair the damaged DNA.

Double-strand breaks (DSBs), like other DNA damage, are a common occurrence in all organisms, but are of particular concern because of the potential for creation of chromosomal aberrations leading to cell death (20). They occur as a direct consequence of exposure of the DNA to ionizing radiation (e.g. x-ray radiation, gamma radiation and cosmic radiation), genotoxic chemicals and metabolic by-products. They also occur as a normal process of immune system development during V(D)J recombination and somatic hypermutation of the immunoglobulin genes (21).

When DSBs occur, they are repaired through one of two pathways, homologous recombination (HR) or non-homologous end-joining (NHEJ). HR repair requires resection of the DNA containing the DSB, followed by strand invasion of a homologous chromosome or sister chromatid by the RAD51-bound single-stranded DNA ends (22). This allows repair synthesis to occur using the homologous DNA as a template. Unlike HR, NHEJ repair does not use a homologous DNA template but instead directly ligates the blunted DNA ends together. As such, this can result in insertions or deletions at the break site and is thus more error prone than the HR repair pathway (22), although a more precise form of the repair has been described (23). Interestingly, components of the

mismatch repair system, which are known to interact with other repair systems, such as NER (24-26), have also been shown to interact with proteins in the HR and NHEJ repair pathways, acting as modulators of recombination and enhancing genetic stability during DSB repair (27-30).

1.2 The DNA Mismatch Repair Pathway

The DNA mismatch repair (MMR) system plays a critical role in genome surveillance and the maintenance of genomic integrity. It is highly conserved among all organisms studied to date (31) and is primarily recognized for its role in the postreplicative repair of DNA damage. The repair system is responsible for identifying and mediating the repair of DNA base-base mismatches that arise during DNA replication; a result of misincorporation of the bases by DNA polymerase. This definition of MMR is somewhat classical in the sense that it is highly oversimplified; we now know that the MMR system proteins are much more diverse with respect to their roles in the DNA damage response and DNA metabolism (6, 7, 32, 33).

Initial identification of the mismatch repair genes was made through the investigation of *Escherichia coli* strains that showed high mutation frequencies (34, 35). These genes were subsequently called mutator (*mut*) genes, of which the *mutS*, *mutL*, *mutH* and *mutU* (*uvrD*) genes are components of the MMR system in *E. coli* (see below). During DNA replication, the DNA polymerase inserts incorrect nucleotides into the nascent strand at a rate of 10^{-4} to 10^{-5} base pairs per replication, which is further reduced to 10^{-7} through the inherent proofreading activity possessed by the polymerase (36). Improper nucleotides

that remain undetected by the polymerase are targets for the MMR system, which further reduces the error rate of DNA replication to approximately 10^{-10}, thus acting to enhance genetic stability.

In the absence of DNA repair, the ability of the cell to repair damaged DNA bases is severely compromised, resulting in the accumulation of spontaneous mutations, including transition and transversion mutations, frameshifts and IDLs (37). Additionally, because of its connections to other DNA repair and signalling pathways, defects in the MMR system can have significant effects on numerous cellular events, including the triggering of apoptosis (38). In humans, aberrant mismatch repair is associated with elevated frequencies of spontaneous mutation (39), microsatellite instability and an increased susceptibility to hereditary non-polyposis colorectal cancer (HNPCC), endometrial and other cancers (40).

1.2.1 Methyl-directed DNA Mismatch Repair in *E. coli*

The MMR pathway has been best characterized in *Escherichia coli*. The main proteins responsible for detection of the mismatch and initiation of repair are MutS, MutL and MutH, of which MutS and MutL are evolutionarily conserved among prokaryotic and eukaryotic organisms (31). Various accessory proteins, including UvrD, one of four exonucleases (ExoI, ExoVII, ExoX or RecJ), single-stranded binding protein (SSB) and DNA polymerase III, are responsible for excision and repair of the mismatch. In *E. coli*, a unique property of this repair system is that it is able to discriminate between the template strand and the newly synthesized daughter strand, to which it directs repair, simply based on the methylation states of the DNA (41, 42). After passage of the DNA

replication fork, there is a delay in the methylation of adenines in the newly synthesized strand by the DNA adenine methyltransferase (Dam), which methylates the DNA at GATC sequences. Thus, repair is directed to the unmethylated strand using the methylated strand as a template for repair of the mismatch.

Mismatch repair is initiated when homodimeric MutS recognizes and binds to a mismatch (Figure 2). This process is dependent on ATP binding and hydrolysis, allowing MutS to dynamically interact with DNA in search for mismatches and subsequently allow its slow release from a bound mismatch (43, 44). Once a mismatch is bound by MutS, a ternary complex is formed consisting of homodimeric MutL, homodimeric MutS and the heteroduplex DNA substrate. Through its ATPase activity, MutL coordinates and regulates MMR by coupling mismatch recognition, by MutS, with downstream activation of other repair proteins. It is for this reason that MutL is often called a "molecular matchmaker" (45). Next, the latent endonuclease MutH cuts the unmethylated strand at a hemimethylated GATC site, 5' to the G. This activity is stimulated in a MutS, MutL and ATP-dependent manner and can occur either 5' or 3' to the mismatch (46). Unlike *E. coli*, MutH is absent in most bacteria. This function is instead covered by MutL (33). Cleavage of the DNA by MutH follows with recruitment of UvrD (DNA helicase II) to the strand break. MutL's ability to bind ATP, not its ability to hydrolyze it, has been shown to stimulate loading of UvrD onto the DNA substrate as well as stimulate the unwinding reaction catalyzed by UvrD (47). Excision of the strand is performed by one of various exonucleases (48), depending on the location of the incision relative to the mismatch. If MutH incises the strand 5' to the mismatch, then excision is performed by

ExoVII or RecJ, of which both contain 5’ to 3’ exonuclease activity. When the incision occurs 3’ to the mismatch, excision is directed by ExoI, ExoVII or ExoX, which contain 3’ to 5’ activity. The resulting gap is stabilized by SSB, filled in by DNA Polymerase III and finally ligated by DNA ligase.

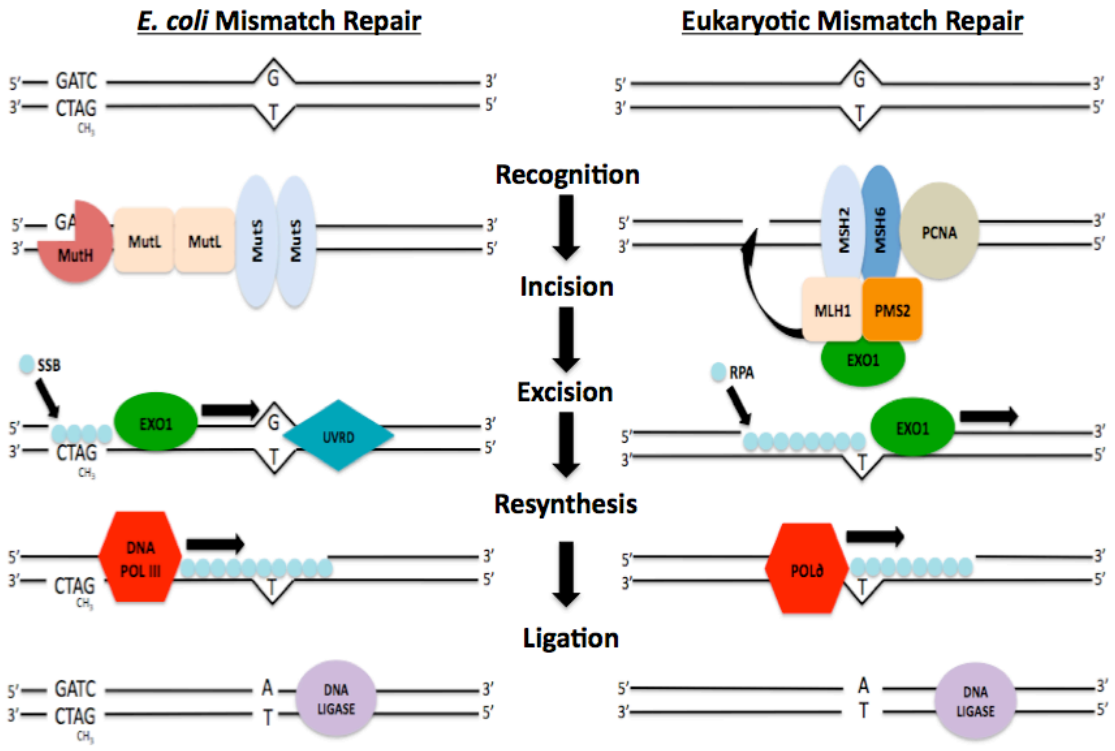


Figure 2. Eukaryotic and prokaryotic post-replicative mismatch repair.

In *E. coli*, MutL mediates mismatch recognition by MutS with strand incision at hemimethylated GATS sequences by MutH. The error-containing strand is excised by one of several exonucleases, followed by binding of SSB and repair synthesis and ligation by DNA polymerase III and DNA ligase I, respectively. In eukaryotes, the MutSα (MSH2-MSH6) heterodimer recognizes DNA mismatches. Interaction with the MutLα (MLH1-PMS2) heterodimer results in incision of the nascent strand either 3’or 5’ to the mismatch. EXOI excises the mismatch containing strand and RPA binds the remaining single-stranded DNA. Repair synthesis and ligation are performed by DNA polymerase δ and DNA ligase, respectively.

1.2.2 DNA Mismatch Repair in Eukaryotes

In eukaryotes, multiple MutL and MutS homologues have been identified but no MutH homologues have been found (Table 1). Homologous genes are related through speciation events (orthologues) or gene duplication (paralogues) (49). Although related, these genes do not always share functional similarities. There are several similarities between mismatch repair in prokaryotes and eukaryotes, including its bidirectionality and basic mechanism, but the presence of multiple MutL and MutS homologues in eukaryotes imparts additional complexity to the system, with many of these homologues having specialized functions. The yeast *Saccharomyces cerevisiae* contains six MutS homologues designated MSH1 to MSH6 (32). MSH2 forms heterodimers with MSH3 and MSH6, termed MutSα and MutSβ respectively, the primary function of which is MMR (50).

Table 1. Components of the prokaryotic (*E. coli*) and eukaryotic mismatch repair systems.

Prokaryotes	Eukaryotes	Function	
MutS	**MutS**α (MSH2 and MSH6) **MutS**β (MSH2 and MSH3), MSH4-MSH5	DNA mismatch recognition IDL recognition Meiotic recombination	**Initiation**
MutL	**MutL**α (MLH1 and PMS2) **MutL**β (MLH1 and PMS1) **MutL**γ (MLH1 and MLH3)	Molecular matchmaker Strand incision (PMS2, MLH3)	
MutH	--	Endonuclease, strand discrimination and incision	**Strand Incision & Excision**
DNA Helicase II (UvrD)	?	Unwinding of DNA	
ExoI, ExoVII, RecJ, ExoX	ExoI	Strand excision	
SSB	RPA	ssDNA protection	**Repair Synthesis**
DNA Polymerase III	DNA Polymerase δ	DNA resynthesis	
DNA ligase	DNA ligase I	DNA ligation	

MutSα can recognize base-base mismatches and insertion/deletion loops (IDLs) of 1 or 2 nucleotides, while MutSβ recognizes larger IDLs (51). MSH4 and MSH5, which interact with each other to form a heterodimer, are involved solely in regulating meiotic recombination and have not been shown to have a role in the repair of base-base mismatches (52, 53). The MSH1 protein, which has been observed in *Saccharomyces cerevisiae* but not mammals, is involved in mitochondrial DNA stability (54).

Eukaryotes also contain two to four MutL homologues, designated MLH or PMS (the prototype of the latter was identified in yeast mutants defective in post-meiotic segregation) (55). These mutants, as well as other MMR mutants, are unable to properly separate chromosomes during meiosis, which can lead to a loss or gain of chromosomes and in yeast a decrease in spore viability (55, 52). In mice, males containing a mutation in the PMS2 gene are infertile, a result of abnormal chromosome synapsis during meiosis, whereas female mice remain fertile (56). In yeast, four MutL homologues have been identified, designated MLH1, MLH2, MLH3 and PMS1, while the corresponding homologues in mammals are MLH1, MLH3, PMS1 and PMS2 (6, 7). The MutLα heterodimeric complex, consisting of MLH1/PMS1 in yeast (57) or MLH1/PMS2 in mammals (58), acts in the MSH2-dependent repair pathway and is responsible for coupling the initial DNA mismatch recognition step to downstream processes. Additionally, two other heterodimeric complexes, designated MutLβ, consisting of MLH1 and PMS1 in mammals, and MutLγ, consisting of MHL1 and MLH3, have been identified. While MutLγ has been shown to function in the repair of single nucleotide insertion/deletion loops (59) and to have a role in meiotic recombination (60), the

function of MutLβ is somewhat unclear, although it's been shown to act in suppressing conservative homologous recombination (61).

Similar to the events in *E. coli* MMR, once a mismatch or IDL is bound by one of the MSH heterodimers, MutLα is able to interact with the complex and trigger downstream repair events, resulting in the excision and repair of the damaged DNA (Figure 2). Numerous additional activities have been identified in mammalian MMR including: the PCNA replication clamp, the RFC clamp loader, DNA polymerase δ, the single-stranded binding protein RPA, exonuclease I (ExoI) and the DNA binding protein HMGB1 (62). While in *E. coli* GATC sites serve as a strand discrimination signal, strand discrimination in eukaryotes is thought to be a consequence of a strand-specific nick; the termini produced during DNA replication (3' terminus on leading strand and 5' terminus on lagging strand) may be the source of these nicks and thus act as strand discrimination signals (51). Furthermore, experiments carried out in the Modrich lab have demonstrated that human PMS2 contains an inherent endonuclease activity that is activated in a mismatch-, MutSα-, RCF-, PCNA- and ATP-dependent manner (63). The incision this complex makes is biased to the discontinuous strand and distal to the mismatch, but can also occur proximal to the mismatch. In the process, a new 5' terminus is formed allowing ExoI, which only exhibits 5' to 3' polarity, to enter and hydrolyze the duplex DNA. DNA polymerase δ is then able to resynthesize the strand and restore its integrity. Interestingly, a recent publication by Josef Jiricny's group has suggested that ribonucleotide monophosphates (rNMPs), inserted into the nascent strand during DNA

replication, can serve as a strand discrimination signal for initiation of the excision step of MMR by EXO1 (64).

1.2.3 Ancillary Functions of the Mismatch Repair Proteins

The DNA mismatch repair proteins in prokaryotic organisms are best known for their functions in DNA repair and recombination. While this was initially the case in eukaryotes, we now know that the mismatch repair proteins in eukaryotes are multifunctional proteins, interacting with a variety of other proteins acting in additional DNA repair and signalling pathway within the cell (Figure 3). One of the most surprising findings is the mismatch promoting role the MMR proteins have in antibody diversity, specifically somatic hypermutation. In order for mammals to efficiently protect against pathogenic threats, the immune system must maintain an incredibly diverse antibody supply. This is accomplished through three pathways, V(D)J recombination, somatic hypermutation, and class-switch recombination (65, 66). During somatic hypermutation, activation-induced cytidine deaminase, which shows a preference for ssDNA, deaminates cytosines (C) to uracil (U) within the variable (V) region of actively transcribed immunoglobulin genes. This results in the formation of U:G pairs and upon binding of the MutSβ heterodimer, removal of the error containing strand. It is at this point that traditional MMR repair synthesis by the error-free DNA polymerases δ and ε is performed by the error-prone DNA polymerases η and κ. This hypermutation within the V region allows for the expansive antibody diversity needed to protect against pathogenic threats.

Another important role of the mismatch repair proteins is the activation of cell-cycle checkpoints and signalling for apoptosis. These two pathways have particular importance to the treatment of cancer with chemotherapeutic agents. Cell cycle arrest occurs when checkpoints associated with the G1, G2 or S phases of the cell cycle are activated. In cells that are exposed to DNA damaging agents, DNA damage that cannot be repaired results in a signalling cascade in which the mismatch repair proteins, or damage sensors, interact with the Chk1 and Chk2 proteins to arrest the cell cycle. Additionally, MMR components can interact with ATM and ATR leading to the induction of apoptosis, through interactions with p53 and p73. This provides one explanation for why cells that are deficient in MMR become resistant to chemotherapeutic drugs. In MMR deficient cells, such as those found in HNPCC, cells are not able to efficiently detect DNA damage, which would normally result in cell cycle arrest and/or apoptosis. Thus cells remain viable and continue to accumulate mutation, a potent trigger of carcinogenesis.

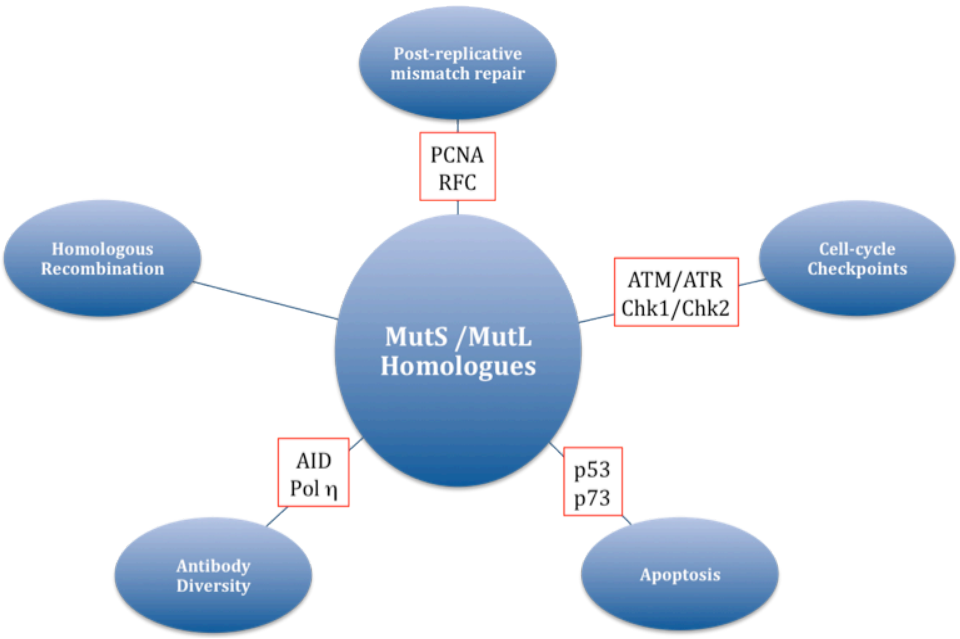


Figure 3. The multifaceted mismatch repair system.
Prokaryotic and eukaryotic MutS and MutL homologues are active in multiple DNA repair and mutation avoidance pathways. Interactions with component proteins in various signalling pathways can result in the activation of cell-cycle checkpoints and triggering of apoptosis. In addition, the MutS and MutL homologues have important functional roles in the recombination of homologous, but non-identical chromosomes and during the recombination events responsible for generating antibody diversity in humans. (adapted from (67))

To further understand the roles of the MMR repair system and its accessory roles in recombination, the cell-cycle checkpoints and apoptosis, we are exploiting the nuclear dimorphism of *Tetrahymena thermophila*. This ciliate protozoan undergoes amitotic, mitotic and meiotic divisions in a single cell, accompanied by extensive genome arrangements and the apoptotic-like process termed programmed nuclear death (PND). *Tetrahymena's* rapid growth in laboratory media, its genetic tractability and large size make it an ideal model organism for the study and analysis of gene function.

1.3 *Tetrahymena thermophila*

1.3.1 History

The history behind the discovery and study of ciliates is an extensive and vast one. It's early beginnings trace back to the 17th century, with the work pioneered by "The Father of Microbiology," Anton van Leeuwenhoek, who observed ciliates, including *Tetrahymena*, during his studies of protists (68). While ciliates have been extensively described and documented since that time, one of the first significant contributions to the area can be attributed to the French microbiologist, André Lwoff, who, in 1923, was able to obtain an axenic culture of the holotrichous ciliate, *Glaucoma pyriformis* GL, which we now know as *Tetrahymena pyriformis* (Erhenberg) Lwoff strain GL, a close relative of *Tetrahymena thermophila* (69, 70).

The importance of *Tetrahymena* as a model organism has been demonstrated by the number of significant discoveries made in the last three decades. These include the elucidation of the structure of telomeres (71) , the discovery and characterization of telomerase (72), the discovery of catalytic RNAs (ribozymes) (73), the functions of histone acetylation on gene expression (74). Additionally, it has been demonstrated that *Tetrahymena* has potential use as a model organism for the expression and study of parasitic antigens and potential vector for vaccine development (75). In terms of the laboratory, *T. thermophila* is an ideal model organism because of its ease of use. It is one of the fastest growing eukaryotic organisms and can be grown in relatively inexpensive media with doubling times of 1.4 to 2 hours at 30˚C (76). Its large size (~20 μm width by ~50 μm length) allows for cytological analysis to be easily performed (77). Also, the ability to observe genetic and cytological events occurring during mitosis and meiosis in

a single cell, makes this organism an ideal model for the study of the DNA mismatch repair proteins. In addition, the unique biology of *Tetrahymena* and its genetic tractability make it possible to introduce linear and circular DNA into either of its nuclei, using biolistic or electrotransformation procedures, thus facilitating the creation of somatic (macronuclear) and germline (micronuclear) knockouts of the MMR genes (78, 79).

1.3.2 Nuclear Dimorphism in *Tetrahymena*

The unicellular eukaryote *Tetrahymena thermophila* is a ciliated protozoan. As with other species in this group, *T. thermophila* exhibits nuclear dimorphism (Figure 4). In a single cell, there are two nuclei, a micronucleus (MIC) and a macronucleus (MAC). The transcriptionally-silent MIC is diploid and contains five pairs of centromeric chromosomes. It is the germline nucleus, and thus the store of genetic information for progeny arising from conjugation. The MAC, on the other hand, is the somatic nucleus and is transcriptionally active, thus acting as the center of gene expression in the cell. This results in the phenotype of the cell being directly connected to the gene expression of the MAC. While its genome is derived from the MIC, the MAC is functionally and physically distinct from the MIC. A principal distinction is that the MAC is a polyploid nucleus consisting of ~ 250 – 300 acentromeric subchromosomal fragments. These fragments are maintained at approximately 45 – 50 times the DNA amount found in the haploid MIC genome (45-50C), thus making up a MAC genome of 104Mb that codes for 27,000 predicted proteins (80). Additionally, the palindromic extrachromosomal rDNA chromosome is an exception to the 45 ploidy, with ~9,000 copies in the MAC and only 1 copy per haploid MIC genome (81).

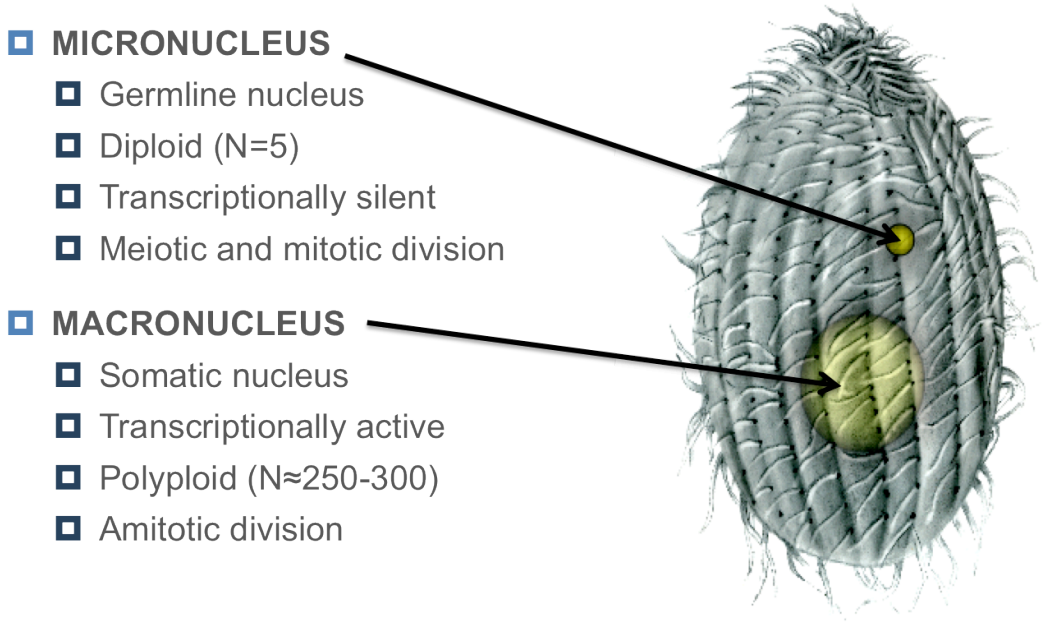


Figure 4. *Tetrahymena thermophila* exhibits nuclear dimorphism.
The transcriptionally-silent micronucleus (MIC) is the germline nucleus and is capable of meiotic and mitotic division. It is a diploid nucleus, containing five pairs of centromeric chromosomes. The transcriptionally-active macronucleus (MAC) provides the phenotype of the cell. It is the somatic nucleus and divides amitotically. It's approximately 250-300 acentromeric subchromosomal fragments are each present at ~45-50 copies per macronucleus and during amitosis randomly assort to daughter cells.

1.3.3 Vegetative Growth and Asexual Reproduction

During the vegetative life cycle of *T. thermophila*, cells reproduce asexually, with the MIC dividing mitotically and the MAC dividing amitotically. Although it is generally assumed that the copy number for each gene in the macronucleus is 45C, chromosomes are randomly distributed to daughter cells during the amitotic division of the MAC (82, 83), a result of the macronuclear sub-chromosomal fragments being acentromeric. Through a process termed phenotypic assortment, successive divisions result in the

initially heterozygous macronuclei becoming homozygous for one allele or another other at all loci (84, 83).

1.3.4 Starvation-induced Conjugation

The prezygotic developmental stage of conjugation occurs after starved cells of two different mating types pair with each other (Figure 5). Each *Tetrahymena* cell can possess one of seven mating types and this is controlled by the *mat* locus. The mating types of progeny cells are determined during conjugation, specifically during the extensive genome rearrangements occurring at postzygotic development, as discussed below. It has recently been shown that mating type is determined through the joining of gene segments located at opposite ends of a tandem array of gene pairs (85). These gene pairs are incomplete, thus upon joining of segments, a fully functional product is formed. This also explains why in cultures that have exited conjugation and began vegetative growth, there is a mixture of mating types.

Following pairing of cells with different mating types, the parental micronuclei proceed through meiosis, resulting in the formation of four haploid pronuclei. Of the four pronuclei, three degenerate, while the remaining prezygotic nucleus undergoes a mitotic division to produce a stationary and migratory pronucleus. The migratory pronuclei are exchanged between partners and fertilization occurs upon fusion with the stationary pronucleus. This restores diploidy and results in a zygotic MIC that undergoes two more mitotic divisions to produce paired cells, each containing two MAC and two MIC. Upon cytokinesis, the process of conjugation has resulted in four progeny cells, each with a

new MIC and new MAC in each cell. The old MAC is degraded in a process termed programmed nuclear death.

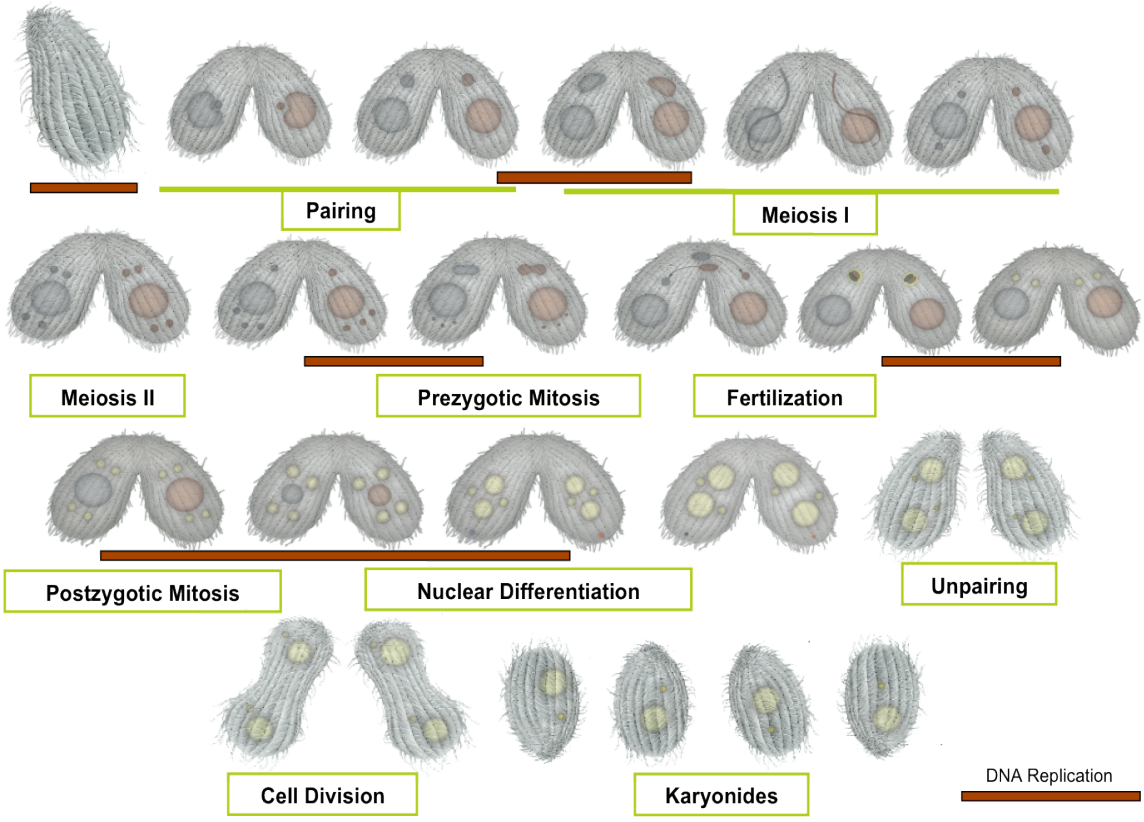


Figure 5. Starvation-induced conjugation in *Tetrahymena*.

Upon mixing of starved cell containing complementary mating types, cells form pairs and proceed through Meiosis I, in which micronuclei elongate and homologous chromosomes recombine. Following meiosis, three of the haploid nuclei degenerate and one is selected to mitotically divide to produce two gametic nuclei, a stationary and migratory pronucleus. Each migratory nucleus is exchanged between mating partners and fertilization occurs upon fusing of the stationary and migratory pronuclei in each cell. This fertilization nucleus, or synkaryons, undergoes several mitotic division and eventually differentiates into two new micronuclei and two new macronuclei. Upon unpairing and feeding, cells undergo cytokinesis and to produce four progeny, or karyonidal, cells. Bars represent periods of DNA replication.

1.3.5 Programmed DNA Rearrangements and DNA Elimination During Conjugation in *Tetrahymena*

Sexual reproduction, in comparison to asexual reproduction, is significantly more complicated in *Tetrahymena*, partly due to the significant number of developmentally programmed DNA rearrangements that occur during the postzygotic development of the nuclei. As discussed above, determination of mating type is one of these rearrangement process. Also included are the fragmentation of the MAC genome into subchromosomal pieces at chromosome break sequences (86) and the removal of ~ 6000 segments of the MIC genome called internally eliminated sequences (IESs), resulting in a MAC genome that is 10-20% smaller than the MIC (80). The sheer number of these IES segments has made it difficult to characterize each one, but there a few that have been relatively well characterized in the past few years; these include the M and R elements (87), Tlr1(88) and mse2.9 (89).

Another event occurring during the second, postzygotic half of conjugation is the apoptotic-like degradation and elimination of the old macronucleus (90). Interestingly, the elimination of this old macronucleus occurs concurrently with the differentiation of the new macronuclei (anlagen), suggesting that this is a programmed process. As such, it has been called programmed nuclear death. It is very similar to apoptotic pathways in other organisms and includes chromatin condensation and DNA fragmentation into nucleosome length fragments. (91).

Chapter 2: Research Objectives, Hypotheses and Rationale

Our current understanding of MMR is biased towards the research that has been performed in a handful of model organisms, including the prokaryote *Escherichia coli* and eukaryotes *Saccharomyces cerevisiae* and *Homo sapiens* (33). Investigation of MMR in these model systems has proven to be very fruitful but fails to provide a clear role for the MMR homologues in development. Additionally, there is still much to be learned about the apparent interactions of the MMR homologues with other proteins in the DNA damage response, cell-cycle checkpoints, apoptosis and genetic recombination.

Tetrahymena thermophila provides an attractive environment in which to study these processes. The nuclear dimorphism of *Tetrahymena* allows for the study of MMR in both a transcriptionally-silent, diploid germline nucleus, that is capable of mitotic and meiotic division, and a transcriptionally-active, polyploid somatic nucleus, that divides by amitosis. The disparity between these two nuclei allows for investigation of MMR under different cellular conditions and in different nuclear environments, all within a single cell. Additionally, *Tetrahymena* provides an opportunity to study the potential roles of the MMR proteins in the massive genome rearrangements and DNA processing events occurring during macronuclear development.

The overarching objective of the research described herein is to characterize the DNA mismatch repair system in the ciliate *Tetrahymena thermophila* and to identify potential functional roles of the MMR proteins during conjugation, with particular attention

focused on events leading up to and including the meiotic development of the micronuclei, prior to postzygotic development of the macronuclear anlagen. The specific objectives of this research are:

1) **Identify putative *MutS* and *MutL* homologues in *Tetrahymena thermophila.***

   Based on the high conservation of the *MutS* and *MutL* homologues in other organisms, we hypothesized that *Tetrahymena* would contain at least the minimal complement of MMR proteins necessary for mismatch repair in a eukaryotic organism, with those being MSH2, MSH6, PMS2 and MLH1. This was investigated using a bioinformatics approach.

2) **Quantify MMR gene transcript abundance in vegetative, starved and conjugating cells.**

   The roles of DNA mismatch repair proteins have been extensively characterized in post-replicative repair, as well as during meiotic recombination and apoptosis. We also know that *Tetrahymena* undergoes several rounds of DNA replication throughout conjugation, a process consisting of meiotic division of the micronuclei and the apoptotic-like programmed nuclear death of the degenerating parental macronucleus, and during macronuclear development. Based on this, we hypothesized that MMR is necessary during conjugation and that this would be reflected as an increase in the transcript levels of the MMR genes. To test this, I used a quantitative real-time PCR based approach using the nucleic acid binding dye, SYTO9, to quantify transcript levels.

3) **Create macronuclear (somatic) knockouts of the MMR genes and analyse associated phenotypes during vegetative and conjugative development by:**

   a. Assessment of morphology and developmental fate of 4',6-diamidino-2-phenylindole (DAPI) stained micro- and macronuclei using a combination of epifluorescence and confocal fluorescence microscopy.
   b. Measurement of nuclear point mutation frequencies in the micronucleus and the macronucleus using a random mutation capture assay.

It is known that cells defective in mismatch repair exhibit a mutator phenotype, in which the frequency of transition and frameshift mutations is increased. In eukaryotic organisms, aberrant mismatch repair is also characterised by microsatellite instability and in male mice, infertility, due to abnormal meiosis (56). We therefore hypothesized that in the absence of functional mismatch repair, *Tetrahymena* would exhibit a mutator phenotype and that this may reveal itself as defects in vegetative growth and/or nuclear development during conjugation, with an accompanying increase in mutation frequency. We specifically hypothesize that MMR is essential in the micronucleus, but not the macronucleus, and that defects in mismatch repair will be seen as an increase in mutation frequency within the MIC , but not the MAC.

The polyploid MAC of *Tetrahymena* divides amitotically during vegetative growth and through the process of phenotypic assortment, the random distribution of alleles eventually results in MAC chromosomes that are homozygous at all loci. We propose that a mutation at single locus of one of the 45 copies of a gene could, through phenotypic assortment, eventually be lost if potentially harmful, or selected for if beneficial, thus negating the need for repair in the MAC. The MIC, on the other hand, has no such mechanism to remove potentially lethal mutations. Mutations that do occur in the MIC will have no effect on vegetative cells, but they can be passed along to progeny during conjugation. If MMR is restricted to the MIC, that nucleus would be expected to have a lower mutation frequency compared to the MAC in wild-type cells. In MMR deficient cells, the MIC should match the higher mutation frequency in the MAC, due to the lack of alternate mutation avoidance mechanism.

In *E. coli*, mutation frequencies can be easily quantified using phenotypic assays such as the Lac reversion assay (92, 93). This entails using strains containing a mutation in the *lacZ* gene, which render it non-functional. In the presence of functional DNA repair, only about 10^{-8} cells becomes Lac^{+}, but in populations of cells with defects in mismatch repair reversion to Lac+ occurs several orders of magnitude more frequently. Unfortunately, using a phenotypic assay to measure mutation rates in the MAC is not possible as there are currently no known phenotypic markers in the MAC. Additionally, the excess

of wild type alleles in this polyploid nucleus would mask the effect of a single mutant allele. Phenotypic assays involving the MIC of *Tetrahymena* are not possible because the MIC is transcriptionally silent.

In 2005, Bielas and Loeb published a genotypic assay, called the Random Mutation Capture (RMC) assay that measures the rate of spontaneous mutation at a single locus within a population of cells (94). It is based on the enrichment and quantification of a mutational target within a unique restriction endonuclease (RE) recognition site that is resistant to digestion due to mutational inactivation of the site. Since the mutated sequence, in contrast to the wild-type sequence, remains intact (undigested), qPCR can be performed to estimate the frequency of mutation. Wild-type sequences that have been digested are unable to serve as templates in the PCR reaction. More recently, in 2011 (95), this technique was simplified through the removal of the technically difficult and impractical enrichment step although if optimized, may be beneficial for use with *Tetrahymena* (see Discussion). Briefly, the simplified RMC assay involves isolation and extensive *Taq*αI digestion of the genomic DNA from a population of cells (Figure 6). Digestion by *Taq*αI occurs at 5'-TCGA-3'sites. In wild-type sequences, digestion is able to proceed, but a mutation in this site renders it resistant to cleavage by *Taq*αI. Thus, the resultant population of DNA molecules will contain digested wild-type and mutant molecules, which than can be quantified using quantitative real-time

PCR. By determining the total number of mutants and total number of base pairs screened, one can determine mutation frequency at a locus of interest.

A key difference between using this assay in mammalian cells versus *T. thermophila* is that one needs to differentiate mutation frequencies in two nuclei, rather than one. To avoid purifying the nuclei separately, which is a difficult technical process, we chose to use MIC- limited sequences and MAC sequences. After mating, the MAC DNA undergoes significant rearrangement, involving deletions of micronuclear-specific internally eliminated sequences (IESs). It is thus possible to use any one of these sequences to measure random mutation in the micronucleus. One such sequence is the mse2.9 IES, located in the second intron of the *ARP1* gene, an acidic protein with unknown function (96, 89). An additional benefit of using this gene is that expression of the *ARP1* gene is not essential for growth of *Tetrahymena*, therefore mutations in this gene should be neutral and not effect the outcome of the assay. In order to measure mutation in the macronucleus any gene or location in the genome, be it intron, exon or untranslated region can be used. We chose the granule lattice protein gene, *GRL1*, which encodes a secretory granule protein, but is not essential for growth (97, 98).

Through a combination of bioinformatics and molecular biology techniques, we have been able to begin to characterize the MMR proteins in *Tetrahymena*. It is hoped that, through this and future research, the complicated biology of *Tetrahymena* can be

understood more deeply, and lead to a better understanding of the roles of mismatch repair in the mutation avoidance and DNA damage pathways as well as in cellular development.

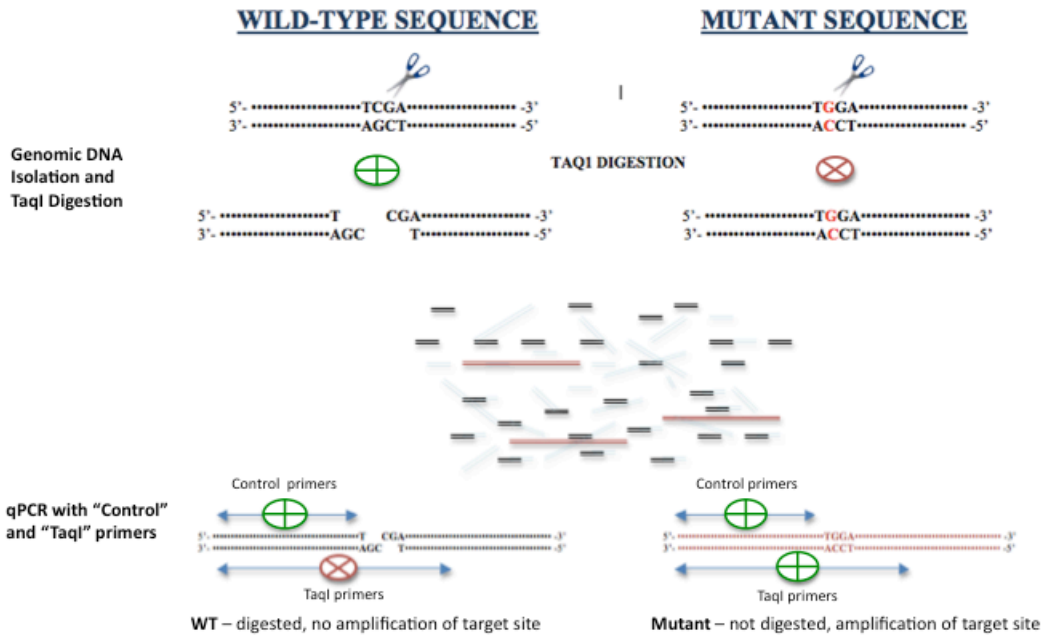


Figure 6. The random mutation capture (RMC) assay.

Genomic DNA (gDNA) is isolated from a chosen sample and is subjected to extensive digestion by the restriction endonuclease *Taq*αI, which cleaves DNA at 5'-TCGA -3' sites. In wild-type sequences, cleavage by *Taq*αI occurs, but in sequences in which the *Taq*αI site has destroyed due to mutation, cleavage does not occur. The gDNA, which contains a population of digested wild-type fragments and undigested mutation containing fragments, is subjected to qPCR experiments to determine mutation frequencies at a particular locus containing a *Taq*αI site. Control primers on one side of the *Taq*αI site are used to determine the total number of molecules/bases screened and primers flanking the *Taq*αI site are used to detect and quantify mutations in the *Taq*αI site. Thus by determining the ratio of mutants detected to total number of bases screened, one can determine mutation frequency.

Chapter 3: Materials and Methods

3.1 Growth, starvation and mating

Tetrahymena thermophila strains B2086.2 and CU428.2 (Table 2) were grown in Super Protease Peptone (SPP) medium (2% protease peptone (GIBCO), 0.1% yeast extract, 0.2% glucose, 0.003% sequestrene (Fe-EDTA, Sigma-Aldrich)), supplemented with 250 U/ml penicillin/streptomycin mix (GIBCO) and 2.5 μg/ml amphotericin B (Fungizone, GIBCO). For vegetative cultures, strains were inoculated into SPP medium and grown at 30˚C with shaking (~250 rpm) to a density of 5 x 10^5 cells/ml. For experiments requiring conjugating cells, cells of complementary mating types were grown as above, followed by pelleting of cells at 800*g* for 5 minutes, then washing and resuspension in 25 ml of starvation buffer (10 mM Tris-HCl, pH 7.5). After 18 - 24 hours of incubation at 30˚C, without shaking, the starved cultures were adjusted to a final concentration of 5 x 10^5 cells/ml. To initiate conjugation, equal volumes of each mating type were combined in a 2 litre Erlenmeyer flask and incubated at 30˚C. Cells were checked microscopically three hours after mixing to verify that the pairing efficiencies were at least 85%. To determine cell counts and pairing efficiencies, cells were fixed with a solution of 5% (w/v) iodine/ 70% (v/v) ethanol and counted using an AO Spencer Bright-Line Improved Neubauer hemocytometer and Leica compound microscope. For pairing efficiencies >100 cells were counted at each hourly time point during conjugation.

3.2 cDNA sequence analysis

cDNA sequences for each of the *Tetrahymena* MMR homologues (obtained by Derek C. Bell) were aligned with the predicted genomic and coding sequences from the

2008 *Tetrahymena* genome sequence annotation (80, 99) available from The Institute for Genomic Research (TIGR; now called the J. Craig Venter Institute) using the alignment software MULTALIN ((100), http://multalin.toulouse.inra.fr/multalin). This allowed us to determine the location of exon-intron junctions, as well as differences, if any, between the sequenced and predicted coding regions. Default parameters, with the DNA comparison table, were used for all alignments. To identify conserved domains and motifs among the putative *Tetrahymena* MMR homologues, sequenced cDNAs were conceptually translated and sequences input into the InterPro (101), Pfam (102) and SMART (103) databases.

Table 2. Genotypes and phenotypes of wild-type strains and MMR knockout strains used and created in this study.

Strain	Micronuclear genotype	Macronuclear genotype	Macronuclear phenotype
B2086.2 (b)	Wild type	Wild type	II
CU428.2 (c)	*mpr1-1/mpr1-1*	*MPR1*	mp-s, VII
SSMSH2b1, SSMSH2b4, SSMSH2b6	*MSH2/MSH2*	*msh2-1[Δ::neo]*	pm-r, II
SSMSH2c1, SSMSH2c2, SSMSH2c3	*mpr1-1/mpr1-1; MSH2/MSH2*	*MPR1; msh2-1[Δ::neo]*	mp-s, pm-r, VII
SSMSH6_1b1, SSMSH6_1b6, SSMSH6_1b13	*MSH6_1/MSH6_1*	*msh6_1-1[Δ::neo]*	pm-r, II
SSMSH6_1c1, SSMSH6_1c4, SSMSH6_1c6	*mpr1-1/mpr1-1; MSH6_1/MSH6_1*	*MPR1; msh6_1-1[Δ::neo]*	mp-s, pm-r, VII
SSMSH6_2b1, SSMSH6_2b4, SSMSH6_2b6	*MSH6_2/MSH6_2*	*msh6_2-1[Δ::neo]*	pm-r, II
SSMSH6_2c1, SSMSH6_2c4, SSMSH6_2c4	*mpr1-1/mpr1-1; MSH6_2/MSH6_2*	*MPR1; msh6_2-1[Δ::neo]*	mp-s, pm-r, VII
SSPMS2b1, SSPMS2b4, SSPMS2b6	*PMS2/PMS2*	*pms2-1[Δ::neo]*	pm-r, II
SSPMS2c1, SSPMS2c4, SSPMS2c6	*mpr1-1/mpr1-1; PMS2/PMS2*	*MPR1; pms2-1[Δ::neo]*	mp-s, pm-r, VII
LSTML1b	*TML1/TML1*	*tml1-1[Δ::neo]*	pm-r, II
EAMSH6_4c	*mpr1-1/mpr1-1; MSH6_4/MSH6_4*	*MPR1; msh6_4-1[Δ::neo]*	mp-s/ pm-r, VII

Strain creator: SS – Shawn Salsiccioli, LS – Lin Sun, EA – Erin Annandale; *mpr* – 6-methylpurine (mp) resistance; Macronuclear phenotype: -s, sensitive; -r, resistant. Mating types are designated by roman numerals. *Msh2-1*, *msh6_1-1*, *pms2-1*, *tml1-1* and *msh6_4-1* are mutant loci that have been disrupted by the neomycin cassette (neo), which confers resistance to paromomycin.

3.3 Phylogenetic analysis

Tetrahymena thermophila MutS and MutL protein sequences were obtained from conceptually translated sequenced cDNAs. MutS and MutL protein sequences from representative prokaryotic and eukaryotic species were obtained from the National Center for Biotechnology Information (NCBI) protein database (Accession numbers available in Appendix A and B). When possible, only full-length protein sequences from the NCBI RefSeq database (104) were used for analysis. Protein sizes of the MutS and MutL homologues used for phylogenetic analysis are shown in Appendix D and E, respectively. Putative MutS and MutL homologues from *Ichthyophthirius multifiliis* (Ich) were obtained by performing a TBLASTN search of the Ich genome (http://ich.ciliate.org, (105)) using *Escherichia coli* MutS/MutL and *Homo sapiens* MSH2, MSH6, MLH1 and PMS2 protein sequences as queries. *I. multifiliis* MSH2 was incomplete at the C-terminal end, but like other ciliate MSH2s, did contain domains II to V, including the conserved ATPase motifs; this missing sequence data did not affect the final ClustalX sequence alignments used for phylogenetic tree construction. Putative MSH6 protein sequences belonging to representative protozoans (Appendix C) were obtained by protein BLAST of the NCBI protozoan genomic BLAST database, using the *Homo sapiens* MSH6 protein sequence. All significant hits with an e-value of $<e^{-49}$ were considered as potential MSH6 homologues and further analyzed by searching the UniProtKB database (106) to identify conserved domains characteristic of the MutS proteins. For *Oxytricha trifallax*, putative MSH6 homologues were identified through BLAST analysis of the *Oxytricha* genome database (OxyDB) (http://oxy.ciliate.org, (107)) using the *H. sapiens* MSH6 protein sequence. Only hits with e-values of $<e^{-100}$ were used for further analysis. Of the

five hits obtained, two (Accession EJY82649.1 and EJY75335.1) appeared to be duplicate sequences as determined from sequence alignment, thus only one of them (EJY82649.1) was used for further analysis. For *T. borealis*, *T. elliotti* and *T. malaccensis*, putative MSH6 homologues (Appendix - Supplementary Table 3) were identified by performing a protein BLAST of the *Tetrahymena* Comparative database (www.broadinstitute.org). Searches were done using *H. sapiens* MSH6 and *T. thermophila* MSH6 protein sequences as queries; hits with bit scores of >100 and >1000, respectively, were considered potential MSH6 orthologues.

Individual multiple sequence alignments were performed for each of the MutS and MutL homologue subgroups using the multiple alignment mode in ClustalX version 2.0 (108) and using full-length protein sequences. For every alignment, the iteration parameter was set to 'iterate each alignment step'. The default multiple alignment parameter 'delay divergent sequences' was set at the default of 30%, except for the MSH1p alignment, which gave a better alignment when set at 20%. Individual alignments were inspected to ensure the correct alignment of conserved regions and to verify that sequences were complete. After inspection, individual alignments were combined using the profile alignment mode in ClustalX 2.0. In the case of *S. lycopersicum* MSH7 and *Z. mays* mus3 (MSH6), only partial sequences were available, but because the conserved C-terminal ATPase region was present in both, they were still used for sequence comparison and phylogenetic analysis. The *I. multifiliis* 'MSH6_3'(EGR27216.1) and *M. brevicolis* MSH6 (XP_001749687) sequences obtained

by BLAST analysis appeared to be incomplete but were still used for phylogenetic analysis, as their removal in initial trees had no effect on tree topology.

Inferred evolutionary relationships of the MutS and MutL homologues were depicted through the construction of Neighbor-joining (NJ) trees in MEGA 5.0 (109) using the 'Poisson correction' amino acid substitution model and pairwise deletion of gaps/missing data. The reliability of internal branches was assessed using the bootstrap phylogeny test; 1000 bootstrap replicates were performed.

3.4 Quantitative real-time PCR

3.4.1 RNA isolation and cDNA synthesis

Total RNA was prepared from both vegetative B2086.1 cells and conjugating (B2086.1 x CU428.1) cells. Prior to starvation, 5 ml (2.5 x 10^6 cells) of a 50 ml culture of vegetatively growing cells were centrifuged at ~2500*g* for 10 min and the cells either resuspended in 1 ml TRI Reagent® (Ambion) or processed according to the protocol supplied with the RNeasy® Mini Kit (Qiagen). After starvation and mixing of the remaining B2086.1 and CU428.1 cultures, 5 ml samples were taken at hourly intervals for nine hours and total RNA extracted according to the manufacturers protocol, as described above. To prevent RNase contamination during RNA isolation, all surfaces, including pipettors, were decontaminated with RNase*Zap*® (Ambion). Following isolation, RNA was resuspended in RNase-free water and stored at -80 °C. Traces of contaminating genomic DNA were removed from all RNA samples using the TURBO DNA-*free*™ Kit (Ambion) or on column digestion with RNase-Free DNase in

conjunction with the RNeasy® Mini Kit. RNA concentrations for each prepared sample were determined using UV-spectrophotometry and measurements were performed in triplicate. Sample purity was estimated from 260nm/280nm ratios, ensuring all were ~2.0. In addition, RNA integrity was gauged through formaldehyde-agarose gel electrophoresis. Samples contained two major bands corresponding to the 17S and 26S rRNA. The 26S rRNA band consists of two halves of the 26S rRNA molecule and is the result of hydrolysis at a hidden break within the molecule, especially under denaturing conditions (110). As such, the corresponding band is of lower molecular weight than expected and produces a 26S:17S ratio of 4:1 rather than 2:1. All samples used for qPCR appeared to be intact. Next, 1 µg of total RNA was annealed with 50 ng random hexamers and cDNA was synthesized using the ThermoScript™ RT-PCR System (Invitrogen). cDNA samples were treated with RNaseH to remove complementary RNA and then diluted 20-fold for use in qPCR analysis.

3.4.2 Quantitative PCR with SYTO9 detection

Prior to expression analysis, forward and reverse primer concentrations were optimized for each MMR and control target gene using a primer concentration matrix of 50, 100, 200 and 300 nM primers with 100 ng genomic DNA as a substrate (Table 3). Once optimized, PCR amplification efficiencies for each primer set were determined using a 1 in 5 dilution series of pooled cDNA (Table 3). The PCR conditions for all primer sets were: 8 min. at 95 °C, 40 cycles of 30 s at 95 °C, 1 min. at 60°C, and 45 s at 72 °C, followed by melt curve analysis (1 min. at 95˚C, 30 sec. at 55˚C and 30 sec. at 95˚C). MMR transcript abundance in vegetative and conjugating cells was determined using

comparative real-time PCR on a MX3000P® QPCR System (Stratagene, La Jolla, CA, USA). Each 15 µl PCR reaction contained: 1X PCR Buffer, 200 uM PCR Grade dNTPs (Invitrogen), 1.5 mM $MgCl_2$, forward and reverse primers (AlphaDNA or Integrated DNA Technologies), 125 nM ROX Reference Dye (Invitrogen), 2 uM SYTO9® Green-Fluorescent Nucleic Acid Stain (Molecular Probes®, Life Technologies™) (22), 1.0 Unit Platinum® *Taq* DNA Polymerase (Invitrogen) and 2 µl of a 1 in 20 dilution of cDNA. All data was normalized to the 17S ribosomal RNA gene and RPL21 genes, since transcript levels of both have been found to be stable in vegetative and starving cells, and throughout conjugation (111, 112).

Table 3. Primers used for quantitative PCR studies of the DNA mismatch repair genes in *Tetrahymena thermophila*.

Gene	GenBank accession	Primer name	Primer sequence (5' → 3')	[ ] (nM)	Amplicon size (bp)	Intron Spanning Primers	PCR Efficiency (%)
Msh2	EF015608	3288B	GCTAAATTCTATAAAGTCGCTAGT	300	631	Yes	89
		3288A	AACTTGCACGTTCCATAGCAGG	300			
Msh6_1	HQ191280	3174-A06	CAGAAGTAGCAAATTCAAGCGT	200	343	No	92
		PE3174r	GAACCAGTTTTCTTGACAGGTG	300			
Msh6_2	EF015607	3297-A02	TGCTTTACAATGGCTGATGGTG	300	681	Yes	85
		1173297-1	TATACCAAAGCTGCTTTCTG	300			
Msh6_3	EF015609	1173238-8	ACAAAAGGTACAGCTGCTCTTC	100	440	No	85
		1173238-5	GACGGGTTCTCTTCACTTCTTG	300			
Msh6_4	HQ191281	4545Ex2In2	TCTGGGGCAAAAGACCTGTTGCTA	300	347	Yes	99
		4545Ex3In2	ACGAATTTGGTTTCCTGCTGGGTC	300			
Pms2	EF015605	F504Ex2In2	ACAGACAGTGATGAAATGGCATACC	100	299	Yes	91
		R504Ex3In2	AGCATATTTTTCATAGGAGAGCCAG	300			
Tml1	EF015606	349101-10	AAAGCAGTAGAACGCTCTTACC	200	362	Yes	111
		349361-5	CGAATTTGCTTGGTGGGTTG	300			
17S rRNA	X54512	17SrDNAForward	ATTCCGGAGAAGGAGCCTGAGAAAC	300	347	N/A	94
		17SrDNAReverse	CTAAGTCGAACCAGTGAAGGCCGAT	300			
Rpl21	M37892	RPL21 Forward (3_3)	TCCTTGAAAGACCCGGTACAA	300	119	Yes	92
		RPL21 Reverse (4_3)	GCCTTCGAATTCGGTCTTGAA	300			

3.5 Creation of MMR knockout strains

3.5.1 Creation of MMR gene-replacement vectors

To create gene replacement/knockout (KO) constructs for each of the MMR homologues, ~ 1kb fragments, corresponding to the 5' and 3' flanking regions of each of the MMR genes, were PCR amplified and successively cloned into the p4T2-1 gene replacement vector (kindly provided by Jacek Gaertig, (113), Figure 15). Restriction endonuclease sites, corresponding to those found in p4T2-1, were included in the 5' ends of the primers. This allowed for cloning of the 5' and 3' flanking regions into corresponding sides of the *neo* cassette, which confers resistance to paromomycin in *Tetrahymena* (114, 113). Primer sequences, corresponding RE sites and amplicon lengths are given in Table 4. PCRs were performed with *Taq* Polymerase (Invitrogen) according to the manufacturers protocol. When necessary, PCR conditions were optimized to maximize the amount of amplicon obtained and to give a single species when analyzed by agarose gel electrophoresis. After PCR, samples were treated with exonuclease 1 (EXO1) (Fermentas) to remove any unused primers and ssDNA products. Reactions included 1U/µl EXO1 and were incubated at 37˚C for 1 hour. Products were purified by phenol:chloroform extraction and ethanol precipitation, in preparation for RE digestion. Restriction digestion was carried out according to the manufacturer's (NEB) protocol. When possible, RE sites for each flanking region were chosen such that the corresponding enzymes were compatible in the same RE Buffer. If not possible, digested products underwent another round of phenol:chloroform extraction and ethanol precipitation before secondary digestion. Cloning was performed using T4 DNA Ligase (Invitrogen) according to the manufacturers protocol. In order to prevent plasmid re-

ligation during cloning, digested p4T2-1 was first treated with Shrimp Alkaline Phosphatase (Fermentas) according to the manufacturers protocol, followed by phenol-chloroform extraction and ethanol precipitation. All KO vectors were confirmed by agarose gel electrophoresis of RE digested plasmids

Table 4. Primers and flanking region sizes of fragments used in the construction of the pKO MMR gene-replacement vectors.

Gene	Flanking region (size, bp)[1]	Primer Name (RE sites underlined)	Primer Sequence (5' → 3')
MSH2	5' (1044)	MSH2 5'forward-KPNI	ATGGTACCACCATACTACTTGGCGAGTA
		MSH2 5'reverse-ECORV	TCGATATCGTCTAGAGAATGAGGGACTC
	3' (1131)	MSH2 3'forward-BAMHI	ATGGATCCGGTCGTGGTACTTCTACAAG
		MSH2 3'reverse-NOTI	ACTGCGGCCGCCTCCTAAGATGTGAATA
MSH6_1	5' (1029)	MSH6_1 5'forward-KPNI	TAGGTACCATATCATAAGCTTTAGTGTG
		MSH6_1 5'reverse-ECORV	ACGATATCCTTCTGTATCATCACAAGTA
	3' (1078)	MSH6_1 3'forward-BAMHI	TAGGATCCATGTTATGTACCTGCTAAAT
		MSH6_1 3'reverse-SACI	ATGAGCTCGAGTGCATCAATGCTTGTAT
MSH6_2	5' (1208)	F5ECORV3297	AGGATATCATCTTCATCGAGTGTACTAAAC
		F5KPNI3297	ATGGTACCAAGATCAATCGATGATCACCCT
	3' (1155)	F3SACI3297	ATGAGCTCGTAGAATAGTACTTGCAAGC
		R3BAMHI3297	ACGGATCCGCATATTTATAGGGCATGAC
MSH6_3	5' (1132)	F5KPNI3238	ATGGTACCAAAGTTTAATCAGACGCTTCAG
		R5CLAI3238	AGATCGATTCTGGCTCATCACTTGGCTAGG
	3' (1219)	F3BAMHI3238	ATGGATCCGCGATGGATAAGTTGAGCAGTC
		R3SACI3238	ACGAGCTCTGGCTACTTCTTAGGATCAAGC
MSH6_4	5' (1019)	F5KPNI3545	CGGGGTACCAGCTGCACAAGCTTAATGA
		R5CLAI3545	CCATCGATGCTCTTCAGCTGGGAGATCT
	3' (1028)	F3BAMHI3545	GCGGATCCGACTATTTGATCAGCCAAG
		R3SACI3545	CGAGCTCAGCATATCATAGCATTAAACTCC
TML1	5' (1148)	F5KPNI8465	GCTGGTACCATTTACAGAACATGGTGAATGC
		R5CLAI8465	GAATCGATGGAGCTTCTTGATCTGAGGACC
	3' (1279)	F3BAMHI8465	ATGGATCCGACATTGTATGAGTTGGATGGCA
		R3SACI8465	ACGAGCTCTTTACCTGACAATCCAGAAAGC
PMS2	5' (1001)	5FKPNI	CGGGGTACCGAATACGTCATATTTTAA
		5RXHOI	CCGCTCGAGTTAGATATGAAAACTCTG
	3' (1027)	3FBAMHI	CGCGGATCCAGTCAAAAAGACATATGA
		3RSACI	CGCGAGCTCAATTTCTCTTAATAAATAAGG

3.5.2 Biolistic transformation

Prior to transformation of *Tetrahymena*, completed MMR gene-replacement vectors were digested with KPN1 and SACI restriction enzymes (NEB) to linearize the plasmid and release the insert. Digests were purified using phenol:chloroform extraction and ethanol precipitation. At least 4 μg of prepared vector DNA was used to coat 40 μl of biolistic gold particles (60 mg/ml prepared) (Bio-Rad) to be used for transformation of macronuclei. The protocol that was used for gold bead preparation and biolistic transformation is detailed in Methods in Cell Biology, Volume 62 (115). Briefly, after starvation overnight at 30˚C, 1.0 x 10^7 B2086.2 and CU48.2 cells were concentrated into 1 ml aliquots and spread onto pre-wet, sterile circular filter paper in a sterile Petri dish. The dish was placed into the bombardment chamber of the BioRad PDS-1000/He biolistic system and cells bombarded with 40 μl of DNA-coated gold beads. Bombardment occurred at a rupture disc burst pressure of 900 psi within a vacuum of 25 psi. Following bombardment, the contents of the Petri dish (including the filter paper) were placed into 50 ml of prewarmed 1% SPP media, incubated at room temperature for ~1 hour, and then incubated at 30˚C. After 6-8 hours, paromomycin (100 μg/ml) was added and cultures aliquoted (100 – 150 μl) into 96-well microtiter plates. Plates were incubated at 30˚C for ~3 days until paromomycin resistant blooms were visible. Those wells showing resistance to paromomycin were subsequently grown with increasing concentrations of paromomycin (100 μg/ml to 2000 μg/ml) to allow for phenotypic assortment. When the growth rate of cultures slowed, single cells were isolated from each well and placed into 2% SPP without paromomycin. After a few days of growth, cells were then placed back into 500 μg/ml paromomycin to ensure isolates were truly

resistant. All manipulations from here on out were done in parallel with all cultures of the same strain (e.g. passaging of cells). Knockouts were confirmed by RT-PCR (SuperScript™ One-Step RT-PCR with Paltinum® *Taq*, Invitrogen), using purified cytoplasmic RNA from 1 x 10^6 cells (RNeasy® Mini Kit, Supplementary Protocol, Qiagen) and primers that were internal to the corresponding MMR gene (Table 3). After selection, knockout strains were kept in soybean culture, as described in (116), in order minimize the number of replications between experiments and avoid the aging of the micronucleus that would normally occur during continual growth and culturing in rich media.

3.5.3 Phenotypic analysis of macronuclear MMR knockout strains during conjugation.

Strains of complementary mating types were grown in 2% SPP overnight at 30˚C. For each of the *MSH2*, *MSH6_1*, *MSH6_2* and *PMS2* KO strains, three strains of the B2086.2 background mutants and three strains of the CU428.2 background mutants were grown up together alongside wild-type B2086.2 and CU428.2, thus representing three biological replicates (Table 2). The next day, cells were counted and adjusted to 3 – 5 x 10^5 cells/ml. Occasionally, strains needed two days to achieve this concentration. In this case, any strains that were approximately at the right concentration were regrown to avoid overgrowth and limit cell death within the culture. Once cultures were at the correct concentrations, they were spun down at 900 *g*, washed with starvation buffer (10mMTris-HCl (pH7.4)) and then resuspended in starvation buffer. After 5 hours, cells numbers were recounted and the concentration adjusted to approximately 2.5 x 10^5 cells/ml. After

at least 16 hours of starvation, 5 x 10^6 cells of a B background strain and its mating C background partner were mixed and incubated at 30˚C, without shaking, to initiate the conjugation process; this was considered time = 0 hours. Pairing efficiencies were determined at every hour. If the wild-type mating was not at least 80% at three hours, matings were aborted and set up again.

To test for completion of mating, a 100 µl aliquot of cells was sampled at 30 hours post-mixing, added to 100 µl of 2% SPP media and grown overnight at 30˚C. The next day, 10 µl samples were added to wells containing 200 µl of 2% SPP supplemented with 15 µg/ml 6-methylpurine (6-mp) or 200 µg/ml paromomycin (pm) and grown overnight at 30˚C. Cultures that grew in 6-mp contained true exconjugants and had successfully completed mating. These were tested for retention of the MAC by growth in 200 µg/ml paromomycin.

3.5.4 Fluorescence microscopy

Cells were either fixed or dried on slides prior to microscopy. For fixation, 1 ml samples were taken every hour throughout conjugation and fixed with 3.3 µl of Schaudinn's fixative (1:2 ratio of 95% ethanol: saturated $HgCl_2$ and 5% glacial acetic acid). After washing with methanol, samples were stored in 1X PBS (pH8.0) until ready for microscopy. Vegetative cells were prepared similarly, except with a primary wash of 10 mM Tris-HCl (pH7.4). Prior to fluorescence microscopy, fixed cells were centrifuged in a tabletop centrifuge, the supernatant removed, and cells resuspended in 20 µl of 1,4-Diazabicyclo[2.2.2]octane (DABCO, Sigma) anti-bleach mounting medium (2.5%

DABCO in 9:1 glycerol:PBS, pH8) containing 4',6-diamidin-2-phenylindole (DAPI, Molecular Probes) at a final concentration of 0.1 µg/ml. Five microliters was spotted onto slides and covered with a coverslip and sealed with clear nail polish. Cells that were dried on slides were stained with DAPI as follows: Slides containing dried cells were prepared and stained according to the published protocol in Methods in Cell Biology, Volume 62 (117).

3.6 Random Mutation Capture Assay

3.6.1 Strains and culture conditions

Strains used for the RMC assay were B2086.2 (wild-type), SSMSH2b1, SSMSH6_1b1, SSMSH6_2b1, EAMSH6_4c, SSPMS2b1 and LSTML1b, as described in Table 2. All strains were maintained in 1% SPP medium (1% proteose peptone, 0.1% yeast extract, 0.2% glucose, 0.003% sequestrene (Fe-EDTA), supplemented with 250 µg/ml penicillin/streptomycin mix (GIBCO) and 2.5 µg/ml amphotericin B (Fungizone, GIBCO). Growth of cells for genomic DNA isolations was done in 30 ml of 2% SPP (same as above except with 2% proteose peptone) at 30˚C with shaking at 250-300 rpm. Cultures were grown for 18 to 48 hours until a cell density of 3.0×10^5 cells/ml or greater was achieved, as determined using a hemocytometer. Cells were spun down at 800*g* for 5 minutes, washed once in sterile 10mM Tris-HCl (pH 7.4), and resuspended in an appropriate volume of 10mM Tris-HCl (pH 7.4) to give an approximate cell concentration of ~ 3.0×10^5 cells/ml. Cells were then starved for 18-24 hours at 30˚C.

3.6.2 Total genomic DNA (gDNA) isolation

The method of genomic DNA isolation used was based on that of Gaertig (113) with slight modifications. After starvation, cell concentrations were checked and, if

necessary, adjusted to 3.0 x 10^5 cells/ml with 10mM Tris-HCl (pH 7.4). For each strain, ~ 1.0 x 10^7 cells were spun down at 800*g* for 5 minutes. All but 500 µl of the supernatant was removed and cells were gently resuspended to break up the cell pellet prior to cell lysis. Forty microliters of Proteinase K (20 mg/ml) was added to the resuspended cells, followed by 3.5 ml of urea lysis buffer (42% (w/v) urea, 0.35 M NaCl, 10 mM Tris-HCl (pH 7.4), 10 mM EDTA and 1% SDS). This solution was gently mixed by inversion and extracted twice with an equal volume of phenol/choloroform/isoamyl alcohol (25:24:1, v/v) and once with chloroform/isoamyl alcohol (24:1, v/v). One millilitre of 5 M NaCl was added to the supernatant and precipitation was facilitated with the addition of an equal volume of isopropyl alcohol. DNA was precipitated overnight at 4˚C and either spooled onto a glass rod or collected by centrifugation. Increased yields were generally obtained by centrifugation. The DNA was washed with 70% (v/v) ethanol, dried and resuspended in 500 µl of 1X TE buffer (10 mM Tris-HCl, 1 mM EDTA; pH 8.0). Six microliters of RNase A (10 mg/ml, Invitrogen) was added to the samples and incubated at 55˚C overnight. A second Proteinase K treatment was performed, followed by another phenol/chloroform/isoamyl alcohol extraction, isopropanol precipitation and ethanol wash, as described above. Samples were resuspended in 400 µl of 1X TE buffer (pH 8.0).

3.6.3 *Taq*αI digestion

The digestion protocol follows that described previously (95), with slight modification. 125 µg of genomic DNA was brought up to a total volume of 600 µl with NEBuffer 4 (10X, New England Biolabs), BSA (100X, New England Biolabs) and sterile, distilled H_2O (sdH_2O) (GIBCO). This mixture was incubated for 1 to 2 hours at 4˚C, then digested with 6 µl (600 U) of *Taq*αI restriction endonuclease (100,000 U/ml,

R0149M, New England Biolabs) at 65˚C for 2 hours. An additional 1 µl (100 U) was added hourly for 4 hours until a total of 1,000 U's had been added to the mixture. When hourly additions were not possible, samples were stored at 4˚C until treatment could be continued. Prior to continuation of digestion, samples were spun down in a tabletop centrifuge and transferred to a new tube. This was done to remove any precipitants that may have formed during the 4˚C incubation. In addition, fresh BSA was added to the sample to aid in digestion. After digestion with 10,000 U's of *Taq*αI, all samples were brought up to 1 ml with NEBuffer 4, BSA and sdH_2O and digested with a final volume of 2.5 µl (250 U) of *Taq*αI for an additional 2 hours. Final samples were spun down, transferred to a new tube and stored at 4˚C.

3.6.4 Quantitative real-time PCR

The protocol was followed as published in (95), with slight modification. All reagents used to perform PCR in the RMC assay were aliquoted into single use tubes to prevent carryover contamination. Also, all surfaces were wiped down with a solution of 10% sodium hypochlorite (bleach) and pipettors and empty eppendorf tubes were sterilized under a germicidal UV lamp for 1 to 2 hours. PCR primers and running conditions for all targets are as indicated in Table 5. Each 20 µl reaction contained: 1X PCR Bugger, 200 µM PCR Grade dNTP's (Invitrogen), 1.5 mM $MgCl_2$, 800 nM forward and reverse primers (Integrated DNA Technologies), 125 nM ROX Reference Dye (Invitrogen), 2 µM SYTO9® Green-Fluorescent Nucleic Acid Stain (Molecular Probes ®, Life Technologies™), 1.0 Unit Platinum® *Taq* DNA Polymerase (Invitrogen) and 5 µl of gDNA. Previous attempts at using dUTP and UDG to treat PCR samples was unsuccessful and therefore omitted from our protocol. A standard curve was made for the

determination of DNA/target copy number in each sample. It contained a series of twelve five-fold dilutions using wild-type B2086.2 gDNA that was diluted to a concentration of 150 ng/µl. The linearity of the dilution series was tested by performing a quantitative PCR with SYTO9 detection and using control primers located in the flanking region of the *Taq*αI site . All other experimental steps were performed as described in (95). After quantitative real-time PCR, melt curves were analysed for the macronuclear target in order to determine which wells contained the correct product and which wells to run on a 3% agarose gel for confirmation. This was not possible for the micronuclear target, due to overlap of MSE2.9 melt curves with primer dimer curves. Therefore all micronuclear targets were analysed by agarose gel electrophoresis. Samples containing product and producing a band of the correct size were digested with 1U of *Taq*αI at 65˚C for 15 minutes, followed by incubation at 95˚C to allow for denaturing of the *Taq*αI restriction endonuclease.

Table 5. Primers and conditions used in the Random Mutation Capture (RMC) Assay.

TARGET NAME	PRIMER NAME	SEQUENCE	PRIMER [ ] (nM)	ANNEALING TEMPERATURE (˚C)	AMPLICON SIZE (bp)	PCR EFFICIENCIES
GRL1 (MAC)	TtGRL1*Taq*I/ctrlFw	GCATCATCATGCACACTAACA	800	60	328	96
	TtGRL1ctrlRv	CCAGATCCGATTTCGTCAG				
	TtGRL1*Taq*I/ctrlFw	GCATCATCATGCACACTAACA	800		518	91
	GRL1*Taq*IReverse	GCTCTTGAAGTCAGCAGCAG				
MSE2.9 (MIC)	MSE2.9(left)ctrlFw	CAAATTCTCCACTTACCTCTTAGAC	800	55	136	93
	MSE2.9(left)ctrlRv	CAAAACCTTTTGGGTGGTTC				
	MSE2.9(left)*Taq*IFw	GAGCTTGCTATTTTAAAGATTGTTTTG	800		313	94
	MSE2.9(left)*Taq*IRv	GGAACAAAACCCATGTAATTTATGA				

PCR RUNNING CONDITIONS				
	GRL1 CONTROL AND *TAQ*I		MSE2.9 CONTROL AND *TAQ*I	
	TEMPERATURE	TIME	TEMPERATURE	TIME
INITIAL MELTING	95	10 minutes	95	10 minutes
MELTING	95	30 seconds	95	30 seconds
ANNEALING	60	1 minute	55	1 minute
EXTENSION	72	1 minute	72	1 minute
FINAL EXTENSION	72	5 minutes	72	5 minutes

Chapter 4: Results

Unless otherwise stated, all experiments presented in this chapter were designed and performed by the author of this thesis, Shawn Salsiccioli. Initial identification of the MMR genes in *Tetrahymena* and cDNA sequencing of the *MSH2, MSH6_1, MSH6_2, MSH6_3, TML1* and *PMS2* homologues was carried out by Derek C. Bell. The *MSH6_4* cDNA sequencing was carried out by Erin J. Annandale and Justin Mui. I have subsequently reanalyzed and updated this sequence data (Table 6 and Figure 7). I constructed all of the MMR gene replacement vectors except for pKOMSH6_4, which was created by Erin J. Annandale and Justin Mui.

4.1 *Tetrahymena* MSH Gene and Protein Structure

As an initial step in identifying putative *MutS* homologues (*MSH*) in *Tetrahymena*, we performed TBLASTN searches of the *Tetrahymena* Genome Database (http://www.ciliate.org, (118, 119)) using *E. coli* MutS and *H. sapiens* MSH2 and MSH6 protein sequences. This search revealed seven putative MutS homologues; one had similarity to eukaryotic MSH2 and four to eukaryotic MSH6 (Table 6). We have named these four putative MSH6 homologues MSH6_1, MSH6_2, MSH6_3 and MSH6_4. The remaining two homologues (TTHERM_00857890 and TTHERM_00763040), although sharing some sequence identity to eukaryotic MSH4 and MSH5 proteins, respectively, were excluded from any further analysis because inspection of conceptually translated sequences, using the InterPro database, indicated that their domain structure was not

conserved (see below). No MSH1 or MSH3 homologues have been identified in *Tetrahymena*.

Table 6. Identities and characteristics of the *Tetrahymena thermophila* DNA mismatch repair *MutS* and *MutL* homologues.

T. thermophila homologue	TTHERM_ID[b]	GenBank Accession[c]	cDNA (CDS) (bp)	# amino acids (kD)	BLASTp (e value)
TML1[a]	TTHERM_00127000	EF015606	2271	756 (87)	*H. sapiens* MLH1 (3e-72) *R. norvegicus* MLH1 (1e-68) *M. musculus* MLH1 (2e-73)
PMS2	TTHERM_01109940	EF015605	2841	946 (110)	*H. sapiens* PMS2 (5e-57) *S. pombe* PMS1 (3e-56) *A. thaliana* PMS1 (4e-54)
MSH2	TTHERM_00295920	EF015608	2442	813 (93)	*S. cerevisiae* MSH2 (1e-141) *H. sapiens* MSH2 (1e-141) *P. tetraurelia* MSH2 (0.0)
MSH6_1	TTHERM_00194810	HQ191280	3699	1232 (142)	*H. sapiens* MSH6 (1e-138) *R. norvegicus* MSH6 (1e-133) *M. musculus* MSH6 (1e-129)
MSH6_2	TTHERM_00426230	EF015607	3405	1134 (131)	*H. sapiens* MSH6 (1e-113) *M. musculus* MSH6 (1e-112) *Z. mays* MSH6 (1e-112)
MSH6_3	TTHERM_00142230	EF015609	3774	1257 (147)	*Z. mays* MSH6 (1e-100) *M. musculus* MSH6 (3e-98) *H. sapiens* MSH6 (7e-97)
MSH6_4	TTHERM_00150000	HQ191281	4170	1389 (161)	*H. sapiens* MSH6 (3e-84) *M. musculus* MSH6 (1e-83) *D. rerio* MSH6 (3e-83)

[a] *Tetrahymena MutL homologue 1 (MLH1 homologue)*
[b] TTHERM ID – *Tetrahymena* Genome Database (TGD) Wiki (www.ciliate.org) gene model identifier
[c] Accession numbers of submitted cDNA sequencing results

For the putative MSH2 and MSH6 homologues, various primer pairs were designed to produce amplicons that were representative of the TIGR predicted coding sequences (CDS). PCR amplicons that were generated using these primers and isolated cDNAs were sequenced and the results compared to the *Tetrahymena* macronuclear genomic sequences available from the *Tetrahymena* Genome Database. Comparisons were made using the MULTALIN software (100), thus allowing us to determine the correctness of the gene models made by TIGR (2008 genome annotation, (99)) and either confirm or refute the predicted gene sequence and organization. Inspection of the

sequenced and predicted *MSH* sequences revealed that the *MSH2*, *MSH6_1* and *MSH6_3* homologues contain the same nucleotide sequence and gene organization as predicted by TIGR, whereas the *MSH6_2* and *MSH6_4* coding sequences contain some small differences (Figure 7). These differences are not sequencing errors, but rather differences in annotation of the intron/exon boundaries.

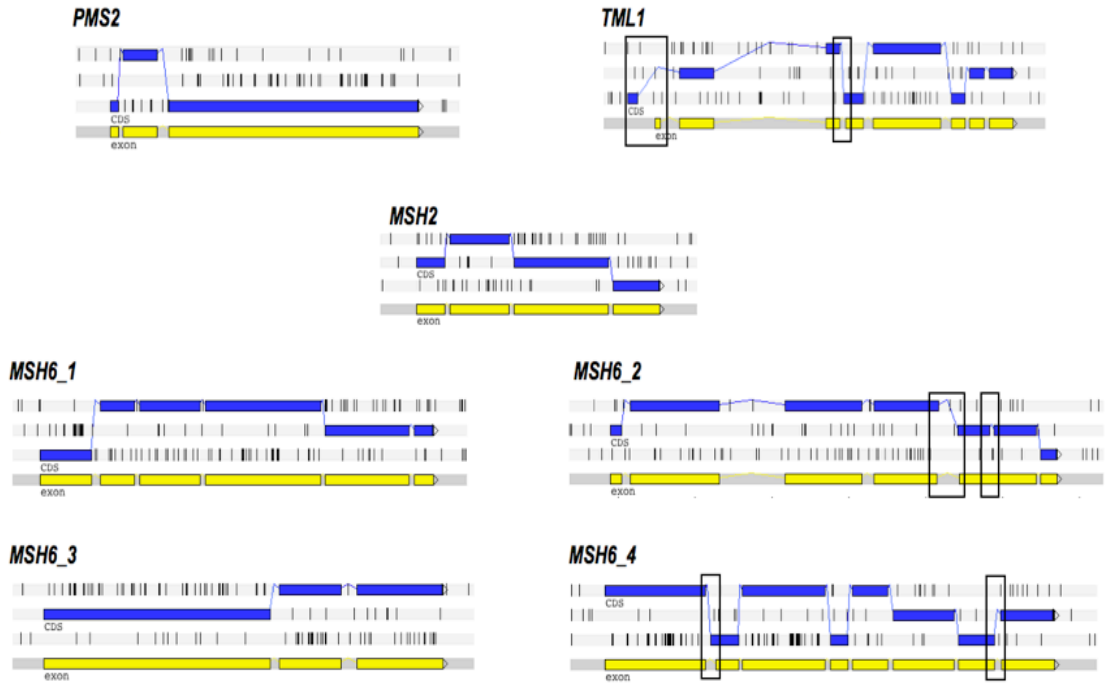


Figure 7. Comparisons of sequenced cDNA and TIGR predicted *MutL* and *MutS* homologue coding sequences (CDS) from *Tetrahymena thermophila.*

Exons of sequenced cDNAs (dark boxes, labelled 'CDS') are shown in their respective reading frames; TIGR predicted exons (light boxes, labelled 'exon'); stop codons within each reading frame are indicated as horizontal black lines. Black boxes indicate differences in gene structure determined from cDNA sequencing and TIGR annotation.

For *MSH6_2*, we found two differences; the first being a difference of 39 bp in the size of intron 4, which was predicted to be slightly larger, and secondly, the presence

of an additional intron in the 3' end of the gene, resulting in 7 exons rather than 6. This results in a CDS size of 3405 bp rather than the predicted 3420 bp. With the *MSH6_4*, some small differences in the starts of exon 2 and exon 8 resulted in a CDS size of 4170 bp compared to the predicted 4107 bp. None of the discrepancies described were found to affect the presence of conserved motifs or domain organizations of the proteins (see below). In addition, all of the introns examined contained the 5'-GT and AG-3' mRNA splice site boundaries, a characteristic of introns from *Tetrahymena* and other eukaryotic organisms (120). When comparing coding sequence GC content of the *MutS* homologues to the GC content of other predicted genes in *Tetrahymena,* the values are in the same range. On average, exons of predicted genes have a GC content of 27.6% (80); *Tetrahymena MSH2*, and *MSH6_1* to *_4* coding sequences have average values of 30.9, 30.5, 29.1, 27.1 and 28.7%, respectively.

MutS proteins, with the exception of the MSH4 and MSH5 homologues, typically contain five conserved domains (121, 122) (Figure 8A). The mismatch binding domain (Domain I) and the clamp domain (Domain IV) are responsible for direct interaction with the mismatched DNA base and for clamping the MutS dimer to the DNA, respectively. In the C-terminus of the protein is the ATP-binding cassette (ABC)-family ATPase (Domain V), and proximal to that, the helix-turn-helix (H-t-H) domain, which is responsible for mediating both MutS dimerization and conformational changes between the MutS subunits in response to ATP hydrolysis (123). Nucleotide binding in the ATPase domain results in conformational changes that are transmitted through the lever domain (Domain III) *directly* to Domain IV and *indirectly* to Domain I via the connector

domain (Domain II), the latter of which has been shown to be important in mediating interactions with downstream proteins, specifically the MutL dimer and MLH1/PMS1 heterodimer (124).

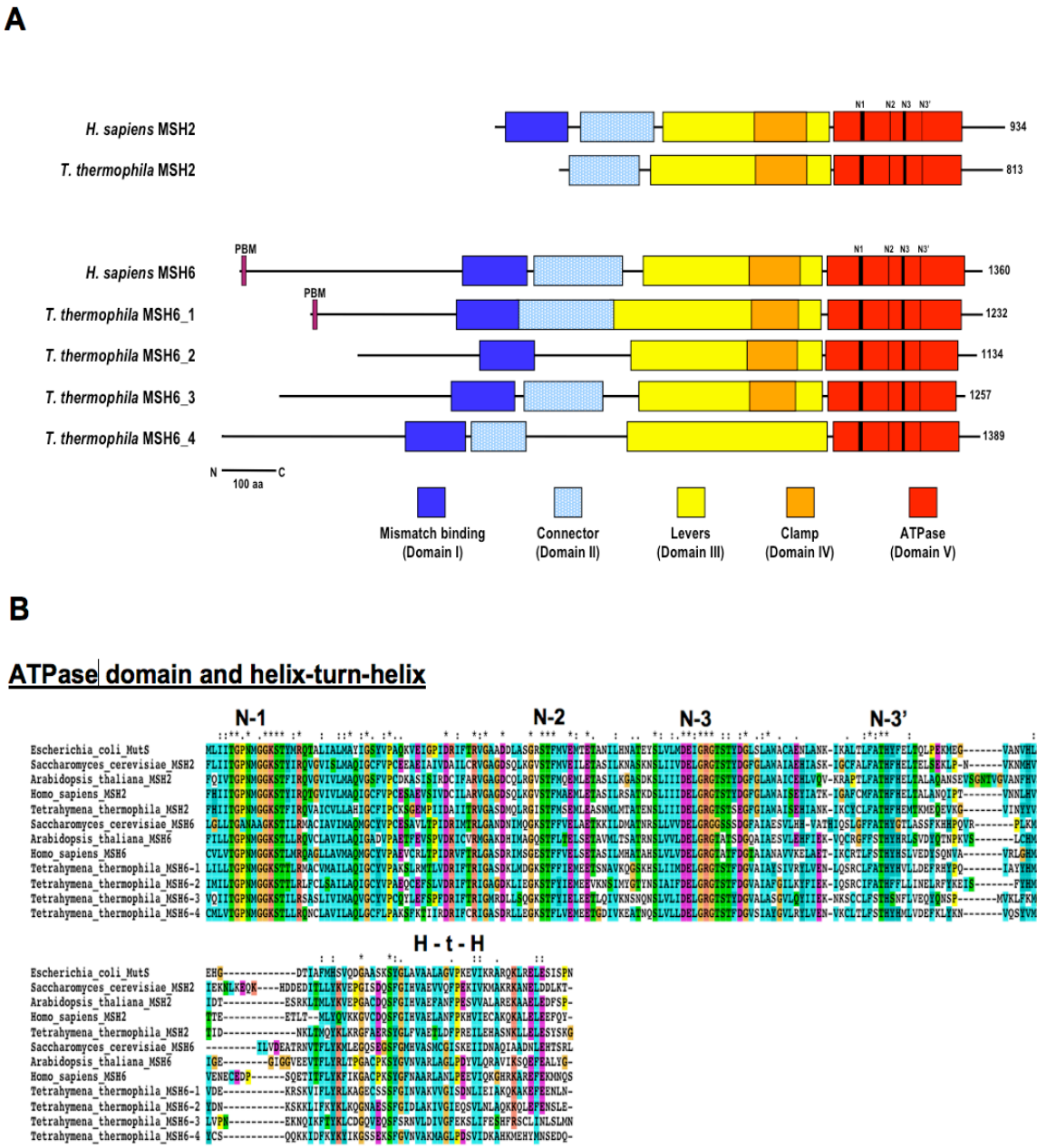


Figure 8. The MSH2 and MSH6 proteins in *Tetrahymena thermophila.*

A. Schematic diagram of the *Tetrahymena* MSH2 and MSH6 proteins, comparing the arrangement of conserved domains relative to those identified in *H. sapiens* MSH2 and MSH6. All representations are shown such that the conserved motifs of the ATPase domain are aligned. Protein sizes are indicated in amino acid residues; Domains I-V are coloured dark blue, light blue, yellow, orange and red, respectively, and functional names are indicated; PCNA binding motif (PBM) is in purple; conserved nucleotide-binding motifs of Domain V are depicted as black horizontal lines and have the following sequences: N1 - Walker A motif (GPNMGGKS), N2 - (ST), N3 – Walker B motif (DELGRG), N3' - (TH). B. Alignment of the conserved C-terminal ATPase and helix-turn-helix domains in prokaryotic and eukaryotic MutS homologues. Positions of the alignment containing identical amino acids are indicated with an asterisk (*), while those positions containing amino acids with highly similar or weakly similar properties are indicated with a colon (:) or period (.), respectively. The locations of conserved nucleotide-binding motifs (N1, N2, N3 and N3') and the Helix-turn-Helix (H-t-H) domain are indicated.

To determine the presence of conserved motifs and domains among the putative MutS homologues in *Tetrahymena*, a search of the InterPro, SMART and Pfam databases was performed (Figure 8A). The *Tetrahymena* MSH2 protein is atypical among the MutS homologues in that it is truncated in the N-terminus and lacks the conserved mismatch-binding domain, resulting in a small, predicted size of 813 amino acids. The absence of this domain implies an inability of the protein to interact with mispaired bases. However, it does not suggest an inability to participate in MMR. In the MSH2/MSH6 heterodimer it is the MSH6 homologue that interacts directly with a mispair (122); in fact MSH2 is dispensable for mismatch binding and repair in yeast (125). Therefore, it is likely that the *Tetrahymena* MSH2 forms a functional heterodimer with one or more of the MSH6 homologues identified. The absence of a mismatch-binding domain, while unusual among eukaryotic MSH2 proteins, may be characteristic of ciliates and some protists, such as *Paramecium tetraurelia, Ichthyophthirius multifiliis* and *Toxoplasma gondii*,

which also appear to possess smaller sized MSH2 proteins and N-terminal truncations of the mismatch binding domain (data not shown).

Of the four MSH6 homologues identified in *Tetrahymena*, only the MSH6_1 and MSH6_3 homologues contain all five of the conserved MutS domains (Figure 8A). The MSH6_2 and MSH6_4 proteins do not have a conserved Domain II and Domain IV, respectively, but this does not necessarily reflect a lack of functional conservation within these regions. Inspection of Domain I in a multiple sequence alignment of representative MSH6 homologues indicates that the *Tetrahymena* MSH6_1, MSH6_2 and MSH6_3 proteins contain the conserved FxE motif. The phenylalanine (Phe, F) and glutamate (Glu, E) within this motif interact directly with mispaired DNA and this binding mechanism is similar among prokaryotes and eukaryotes (121, 122), Thus, it appears as though these three MSH6 homologues may be functional in the repair of single base mismatches, whereas the MSH6_4 may have a role in the repair of insertion/deletion loops or perhaps some other biological role within the cell.

While all the putative *Tetrahymena* MutS homologues contain the highly conserved nucleotide-binding motifs comprising the ATPase domain; N1 (Walker A motif or P-loop, GPNMGGKS), N2 (ST), N3 (Walker B, DELGRG) and N3' (TH) (Figure 8B, (126)), other motifs are present in just a subset of the proteins. A unique feature of eukaryotic MSH3 and MSH6 proteins is the conservation of the proliferating cell nuclear antigen (PCNA) binding motif at their N-terminal end. This motif has the consensus Qxxhxxaa, where 'h' is a hydrophobic amino acid and 'a' is an aromatic acid,

and its interactions with MSH3, MSH6 (127, 128) and MutLα (129, 130) have been shown to be important for functional mismatch repair and genome stability. To determine if this motif is present in the *Tetrahymena* MSH6 proteins, a regular expression search (131) of the MSH6 sequences was performed using the search pattern Q..[AILV]..[FWY][FWY]. Only the MSH6_1 homologue was found to contain the conserved PCNA binding motif (QSSLLSFF, amino acids 3-10), suggesting that it may be the primary MSH6 homologue functioning in DNA mismatch repair. A search for the conserved helix-turn-helix (H-t-H) structural domain revealed that while the MSH2, MSH6_1 and MSH6_4 proteins contain the conserved motif (G..VA.....P), the MSH6_2 and MSH6_3 proteins do not (data not shown). Further analysis of the predicted secondary structure of the proteins, using Scratch Protein Predictor (132), indicates that all of the *Tetrahymena* proteins do contain a H-t-H motif and/or structure near the ATPase domain in their C-terminus, which is also evident in multiple sequence alignments of the MSH sequences (Figure 8B).

4.2 *Tetrahymena* MSH protein phylogeny

The evolutionary relationship of the *Tetrahymena MSH2* and *MSH6* genes with other species was explored through the construction of a distance based Neighbor-Joining (NJ) tree, using an alignment of MSH sequences from representative prokaryotic and eukaryotic species. To maximize the information about these evolutionary relationships, complete sequence data was used for the analysis, rather than using only the highly conserved C-terminal ATPase domain, as also suggested by Culligan *et al*. (133). Use of

the conserved C-terminal ATPase region of the MutS and MSH sequences alone results in low bootstrap support for most branches (personal observation and (133)). The NJ tree (Figure 9), rooted to bacterial MutS homologues, shows eight distinct paralogous (related by gene duplication) groups, MutS and MSH1-MSH6. Plant MSH1 and that of *Toxoplasma gondii* (134) are diverged from fungal MSH1 and thus each forms its own MSH1 subgroup within the tree. Because of this significant divergence, the placement of plant MSH1 was not clearly resolved. Relative to other eukaryotic MSH homologues, the MSH2 and MSH6 homologues of *Tetrahymena* are somewhat diverged. The MSH2 homologue, which is N-terminally truncated relative to other eukaryotic MSH2 proteins, is placed within the MSH2 clade and is most closely related to other protist MSH2 orthologues (related by speciation events). It shares highest similarity with its sister orthologue in the ciliate *I. multifiliis*, both of which are closely related to the *P. tetraurelia* orthologue. The *Tetrahymena* MSH6 homologues are considerably diverged from other eukaryotic MSH6s, forming their own subgroup within the MSH6 clade. They share the most similarity with the *T. gondii* MSH6 orthologue, although this branch was poorly resolved and may not be reflective of a true relationship. Inspection of the *Tetrahymena* Genome Database (119) reveals that the MSH6_2 homologue has been annotated as MSH3, but based on the grouping of this homologue with other MSH6s and not MSH3s, this annotation is incorrect.

A deeper phylogenetic analysis of confirmed and putative MSH6 sequences from a variety of protists and *Tetrahymena* species shows that the presence of multiple MSH6 homologues is common among members of this group, specifically the ciliates (Figure

10). Again, as with the previous NJ analysis, the MSH6 homologues of *Tetrahymena thermophila* and its related species are quite diverged from those of other protists such as *Trypanosoma cruzi* MSH6 and MSH8 and *Toxoplasma gondii* MSH6. Interestingly, BLAST analysis of the recently sequenced *Ichthyophthirius multifiliis* genome revealed three MSH6 homologues. All three of these homologues group individually with separate MSH6s from *Tetrahymena*, either the MSH6_1, MSH6_3 or MSH6_2. In addition, BLAST analysis of the *Paramecium tetraurelia* genome indicated the presence of five MSH6 homologues, although they appear distantly related to the *Tetrahymena* homologues and form two subgroups, one with the *Tetrahymena* MSH6_1/MSH6_2 clades and one with the MSH6_4 clade. We also identified four MSH6 homologues in *Oxytricha trifallax*, although they were less related to any of the *Tetrahymena* homologues and formed their own subgroup within the ciliate MSH6 clade.

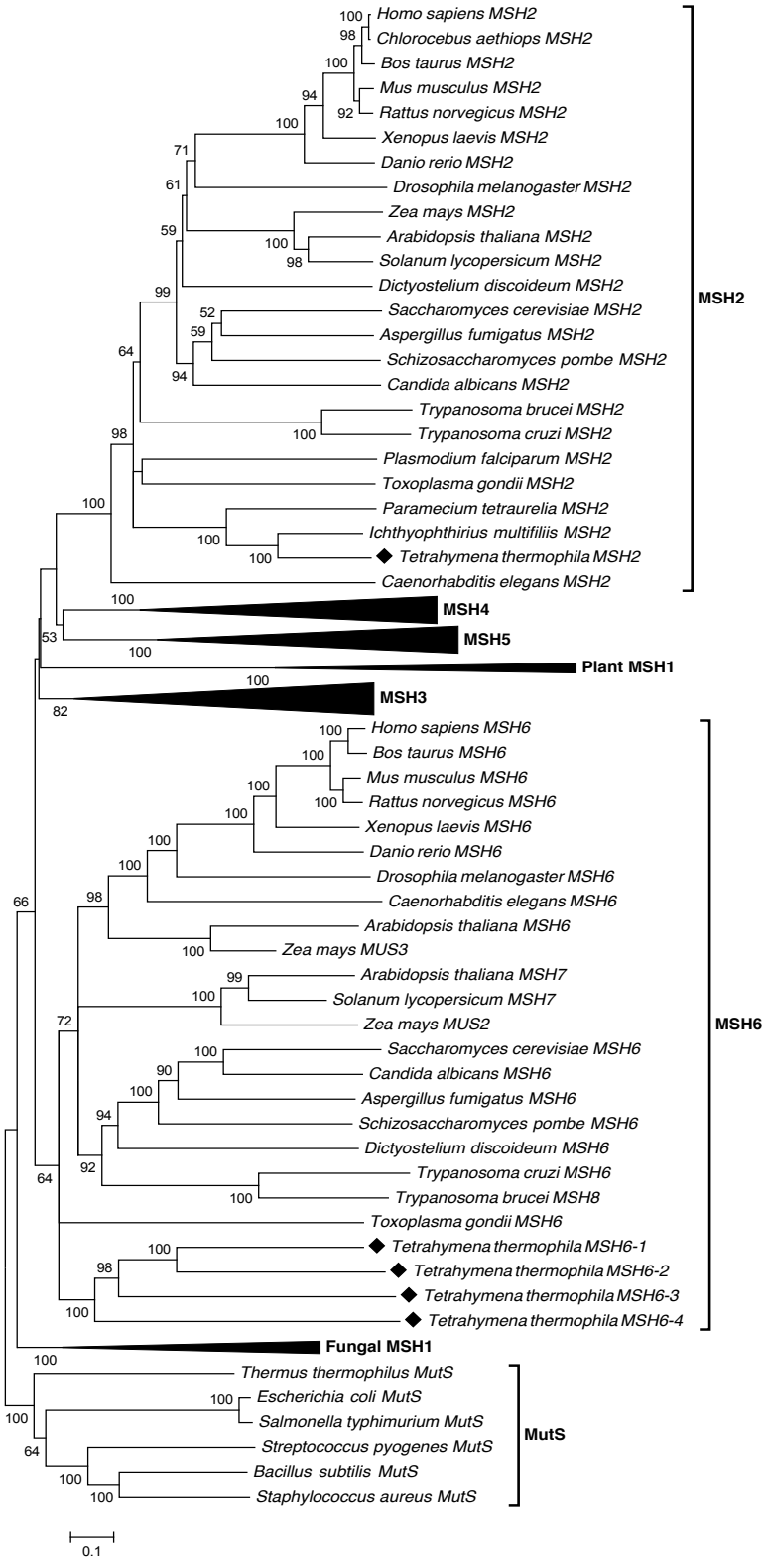


Figure 9. Phylogenetic tree of prokaryotic and eukaryotic MutS homologues.

Phylogenetic tree showing evolutionary relationship of *T. thermophila MSH2*, *MSH6_1*, *MSH6_2*, *MSH6_3* and *MSH6_4* with representative prokaryotic and eukaryotic *MutS* homologues. The tree was rooted using the prokaryotic MutS clade as an outgroup. Evolutionary relationships were inferred using the Neighbour-Joining method. Bootstrap values (1000 replicates) are indicated next to each branch if reproduced in >50% of replicates. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. For clarity, the MSH1, MSH3, MSH4 and MSH5 clades have been collapsed.

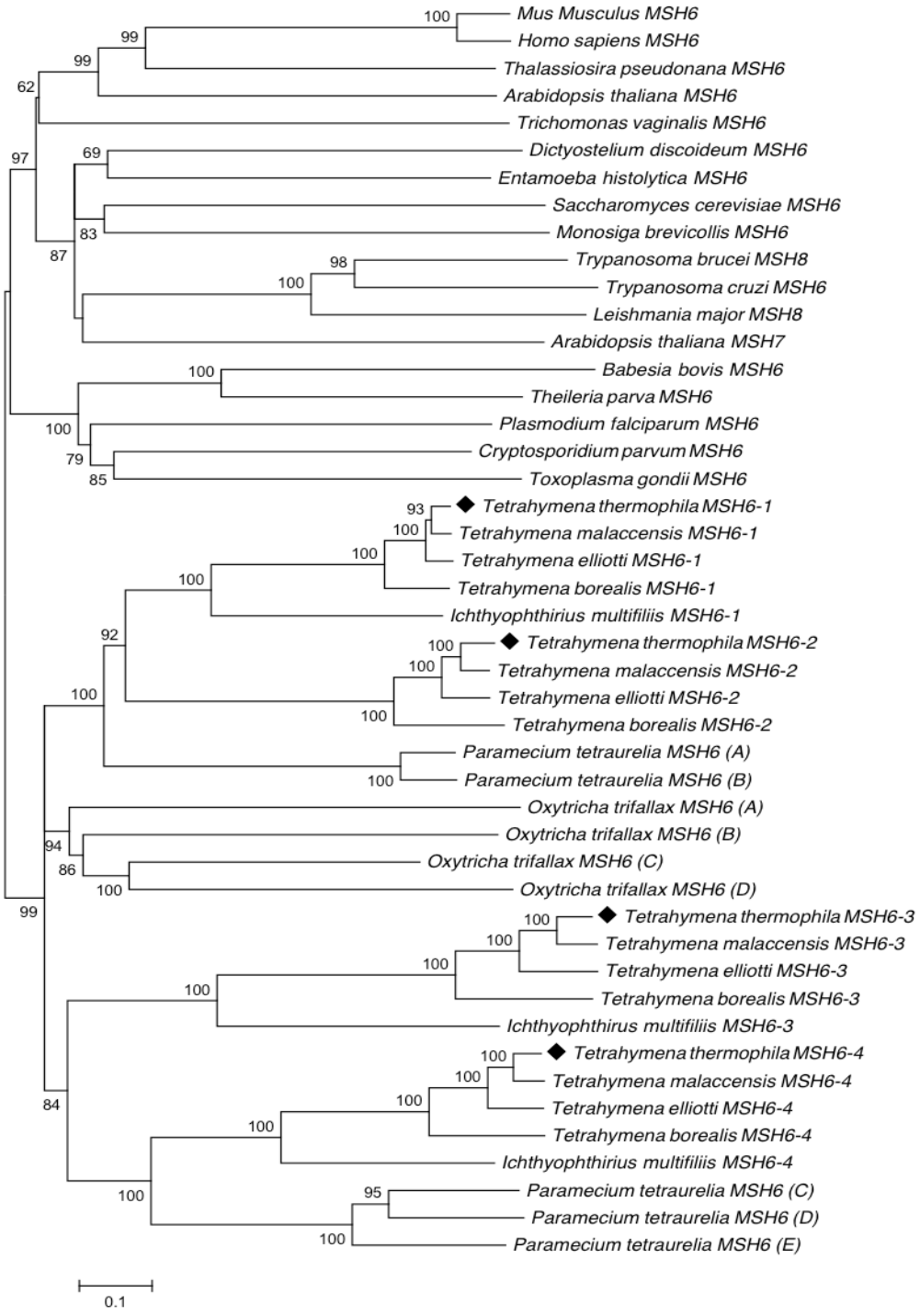


Figure 10. Phylogenetic tree of protozoan MutS homologues.

Unrooted phylogenetic tree showing the evolutionary relationship of *T. thermophila MSH6* homologues with representative protozoan species. Tree construction methods and bootstrap values are as given in Figure 9.

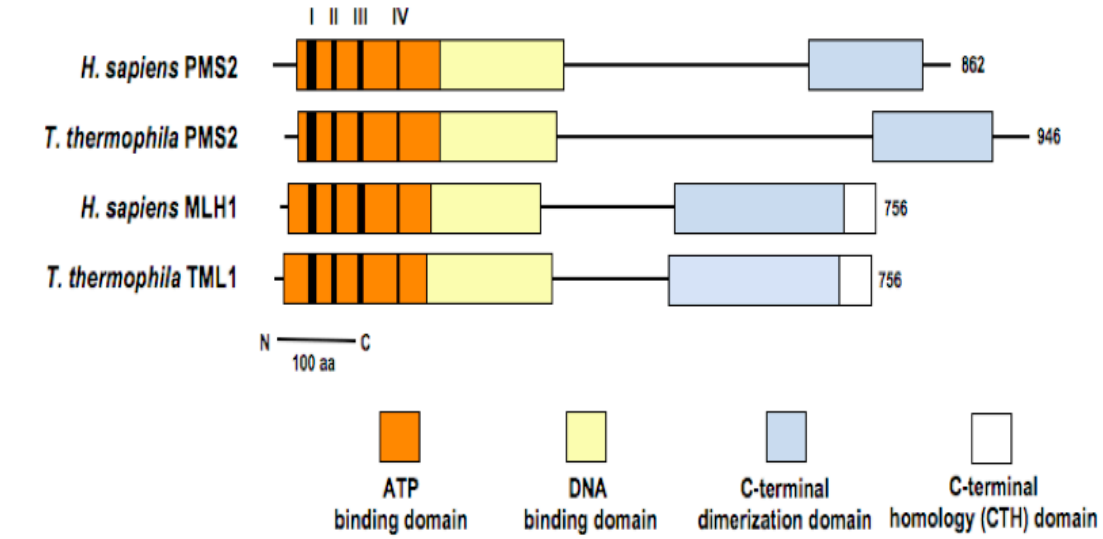
4.3 The MutL homologues of *Tetrahymena thermophila*

A TBLASTN search of the *Tetrahymena* Genome Database using the *E. coli* MutL and *H. sapiens* MLH1 and PMS2 protein sequences revealed two strong MutL homologue (MLH) candidates, one of which contains the metal-binding domain (DQHAxxExxxxE) found in PMS2 and MLH3 homologues (63). A third, MLH3-like candidate contains the metal-binding domain, but lacks the highly conserved ATPase domain characteristic of members of the MutL family and was not analysed further (see below). An initial BLAST search of the NCBI database, using the TIGR predicted sequences, indicated that one of the homologues was most closely related to eukaryotic PMS2 homologues, while the other most closely resembled eukaryotic MLH1 homologues (Table 6). We have thus named the one homologue *PMS2* and the second homologue *TML1* (Tetrahymena MutL), since *Tetrahymena* already contains an *MLH1* gene (Micronuclear Linker Histone 1) (135). cDNA sequencing of these two homologues revealed that the gene model for *PMS2* was correctly predicted by TIGR and that of *TML1* was incorrectly predicted by TIGR; there were small differences in exon/intron boundaries and a significant difference in the location and size of exon 1 (Figure 7). Our cDNA sequencing indicates that exon 1 of *TML1* is 109 bp and it begins 328 bases upstream of the predicted exon 1 (55 bp) location. We also observe that exon 4 is 222 bp, rather than the predicted 201 bp. These differences result in a CDS size of 2271 bp rather than a predicted size of 2196 bp but do not affect the conserved domains or motifs characteristic of the MLH proteins. Additionally, as with the introns of the *MutS* homologues and other *Tetrahymena* genes, the introns of the *MutL* homologues contain the 5'-GT and AG-3' mRNA splice site boundaries (120). The GC content for the coding

sequence of the *PMS2* homologue was 25.5% GC and that of the *TML1* coding sequence was 31.4%, both in range of the predicted average of 27.6% (80).

Conserved domains and motifs characteristic of the MutL family were identified in the putative *Tetrahymena* MutL homologues through searches of the InterPro database using conceptually translated protein sequences derived from our sequenced cDNA. MutL homologues consist of two functional regions: a highly conserved N-terminal ATPase domain and a less well conserved C-terminal dimerization domain (Figure 11A). All MutL family members are part of the GHKL ATPase superfamily, which includes a diverse population of proteins, including those related to Hsp90, DNA topoisomerase II, and bacterial histidine kinases (136). A key feature of GHKL ATPases is the conservation of four motifs in the ATP-binding region, known as the Bergerat fold (137). Examination of both the TML1 and PMS2 proteins of Tetrahymena indicates that these two proteins are functional ATPases as they contain the necessary conserved motifs; Motif I – ExxxNxxDA; Motif II – DxGxG; Motif III - GxxGx[GA]; Motif IV – GT (Figure 11B). The C-terminal dimerization domain, while conserved in the PMS2 homologue in this study, was not well conserved in the TML1 homologue as indicated by sequence comparison using the InterPro database. This is not unusual among members of the MutL family, which tend to have relatively divergent C-terminal sequences, but retain C-terminal structural homology (138)**.** Unique to other eukaryotic MLH1 homologues is the C-terminal homology domain (CTH), consisting of the K[I/V]FERC motif (139). The TML1 protein does contain this motif (Figure 11B).

The evolutionary relationships of *Tetrahymena* TML1 and PMS2 with other prokaryotic and eukaryotic MutL paralogues were inferred from a phylogenetic tree constructed using the NJ method (Figure 12). As with the MutS phylogenetic analysis, full-length sequences were used in order to maximize the information obtained regarding inferred evolutionary relationships (133). The rooted NJ tree shows five distinct paralogous groups, prokaryotic MutL and eukaryotic MLH1, MLH3, PMS1 and PMS2. The *Tetrahymena* MutL homologues are diverged relative to other those of other eukaryotic species. The TML1 protein groups with the eukaryotic MLH1 clade and the PMS2 homologue with the PMS2 clade. The *Tetrahymena* PMS2 protein shares the most similarity with other ciliate PMS2s and resolves with the *P. tetraurelia* and *I. multifiliis* PMS2 orthologues. Similar to the ciliate MSH6's (Figure 9 and Figure 10), the PMS2 homologues of ciliates are found in their own distinct clade, suggesting these homologues are much more divergent than their other protist counterparts. Although the TML1 homologue grouped with other protist MLH1 homologues it did not resolve well, and thus this may not be reflective of a true evolutionary relationship. This may also be the reason for the apparent divergence from the ciliate *P. tetraurelia* MLH1.



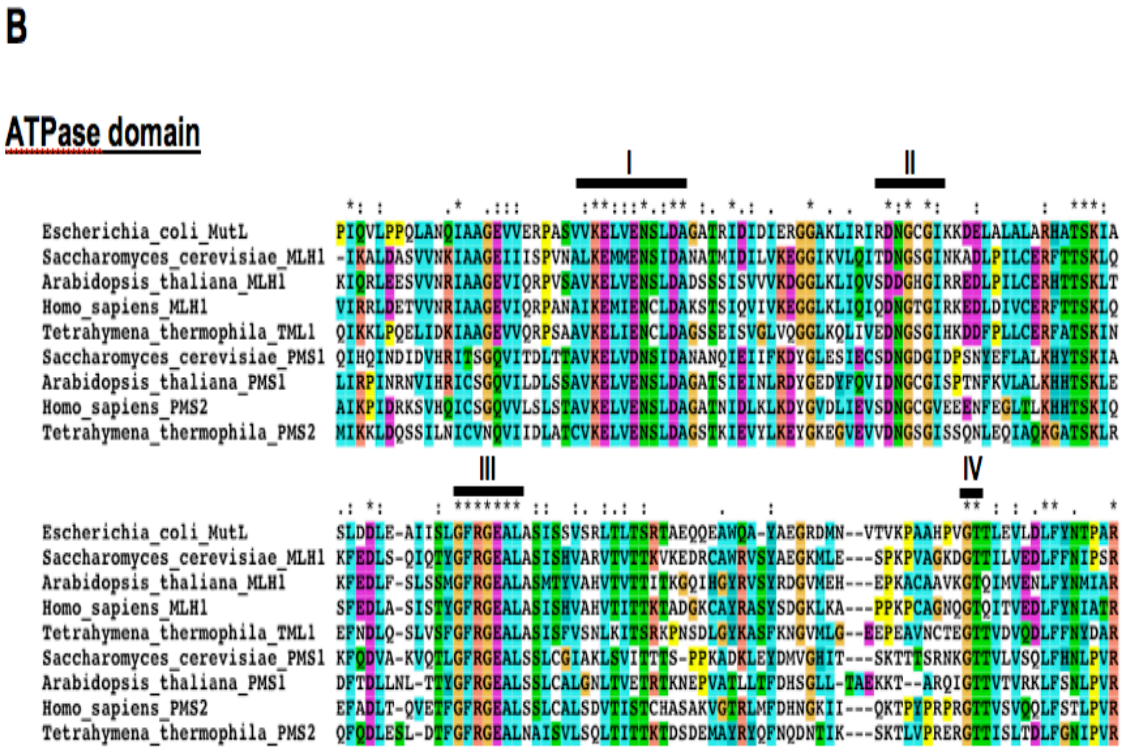


Figure 11. The PMS2 and TML1 proteins in *Tetrahymena thermophila*.

A. Schematic diagram of the *Tetrahymena* PMS2 and TML1 proteins, comparing the arrangement of conserved domains relative to those identified in *H. sapiens* PMS2 and MLH1. All representations are shown such that the conserved nucleotide-binding motifs (I-IV) of the ATPase domain are aligned. Protein sizes are indicated in amino acid residues; ATP-binding /ATPase domain – orange; DNA binding domain – yellow; C-terminal dimerization domain – blue; C-terminal homology domain (CTH) – white; conserved nucleotide-binding motifs of the ATP-binding domain are depicted as black horizontal lines and have the following consensus sequences: Motif I – ExxxNxxDA, Motif II – DxGxG, Motif III - GxxGx[GA], Motif IV – GT. B. Alignment of the conserved N-terminal ATPase domain in prokaryotic and eukaryotic MutL homologues. Positions of the alignment containing identical amino acids are indicated with an asterisk (*), while those positions containing amino acids with highly similar or weakly similar properties are indicated with a colon (:) or period (.), respectively. The locations of conserved nucleotide-binding motifs (I, II, III, and IV) are indicated.

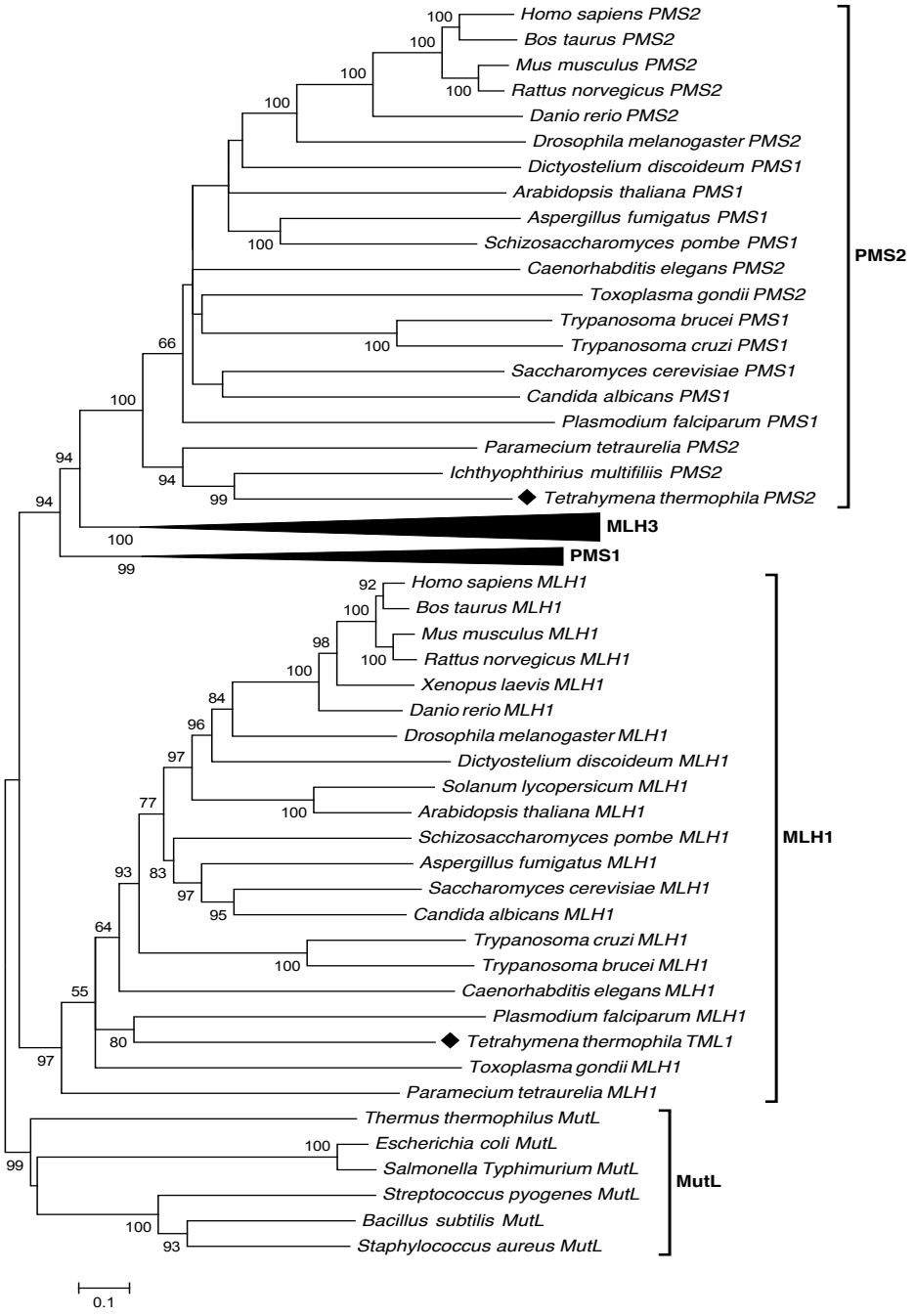


Figure 12. Phylogenetic analysis of the *Tetrahymena thermophila PMS2* and *TML1* homologues. Evolutionary relationships were inferred using the Neighbor-Joining method and using full-length protein sequences. The tree was rooted using the prokaryotic MutL clade as an outgroup. Tree construction methods and bootstrap values are as given in Figure 9. For clarity, the MLH3 and PMS1 clades have been collapsed.

4.4 Lack of meiosis-specific MSH4/ MSH5 and MLH3 in *Tetrahymena*

Protein BLAST analysis and inspection of the *Tetrahymena* Genome Database revealed two additional MutS homologues. A secondary protein BLAST of the NCBI non-redundant protein sequence database indicated one had partial homology to the ATPase domain (Domain V) of eukaryotic MSH5 proteins. Although this predicted protein contained the conserved Walker A (GPNYAGKS) and Walker B (LILIDE) motifs characteristic of MutS homologues (Appendix Supplementary Figure 1), the ATPase domain was located in the N-terminus rather than C-terminus of the protein, completely unlike all other MutS homologues. This suggests that this protein is neither structurally nor functionally similar to other MSH5 proteins. The other putative MutS was similar to eukaryotic MSH4 proteins, but again query coverage and BLAST scores were low. Although this protein is predicted to have a conserved connector domain (Domain II) and lever domain (Domain III), these regions are smaller than those found in other MutS homologues (Appendix F). The predicted ATPase domain does contain the Walker A and Walker B motifs, but as with the other two domains, the predicted size is small compared to other eukaryotic MutS homologues. Further, phylogenetic analysis of both of these homologues shows that they do not group with known MSH4 and MSH5 homologues, and thus are most likely unrelated or at best very distantly related (data not shown). We have also identified a putative MLH3 homologue, but the predicted size of the protein was very small (319 aa) in comparison to other MLH3 homologues, and inspection of the InterPro database indicated that the only conservation was in the C-terminal dimerization domain (amino acids 94-271). Interestingly, this putative protein does contain the metal-binding motif found in PMS2 and MLH3 homologues (63).

4.5 MMR gene transcripts are upregulated during conjugation

We have used quantitative real-time PCR (qPCR) to measure the transcript abundance of each of the MMR homologues during conjugation and vegetative growth. Starved cells represent a point in which there are low levels of DNA replication, transcription and protein synthesis (140-142), so all transcript abundance values during vegetative growth and conjugation have been expressed relative to those determined in starved cells. Standard curves of the MMR gene amplicons, as well those for the *RPL21* and *17S rRNA* normalizer genes, were generated using a five-fold dilution series in order to determine PCR efficiencies (Table 3). Due to the differences in PCR efficiencies among targets, relative quantification of transcript abundance was performed using the method of Pfaffl (143), which employs PCR efficiencies in quantitative calculations in order to correct for differences in amplification among targets, thus allowing for accurate quantification and normalization of expression data. This is the preferred method of amplification when PCR efficiencies have not satisfied the assumption of equality needed for the $2^{-\Delta\Delta Ct}$ method of quantification (144).

To initiate conjugation, two starved wild-type strains of *Tetrahymena* (B2086.1 and CU428.1, Table 3) expressing different mating types were mixed. Samples were taken at hourly intervals and then mRNA transcript levels measured. During conjugation, the abundance of MMR gene transcripts is elevated, relative to starved cells, for all MMR genes (Figure 13). This increase begins when cells pair at around 1 hour and reaches its maximum at 3 to 4 hours after mixing. This observation was consistent across all genes

other than for the *MSH6_3* gene, which has maximal transcript abundance at 5 hours. The period between 3 and 4 hours corresponds to meiotic prophase of Meiosis I, which culminates in the elongation of the micronuclei and pairing and recombination of homologous chromosomes (112), whereas 5 hours corresponds to the period at which the postzygotic mitosis of micronuclei occurs (Figure 14). Vegetative samples had lower levels of transcript than starved cells and may represent the basal level of transcription necessary for vegetative growth.

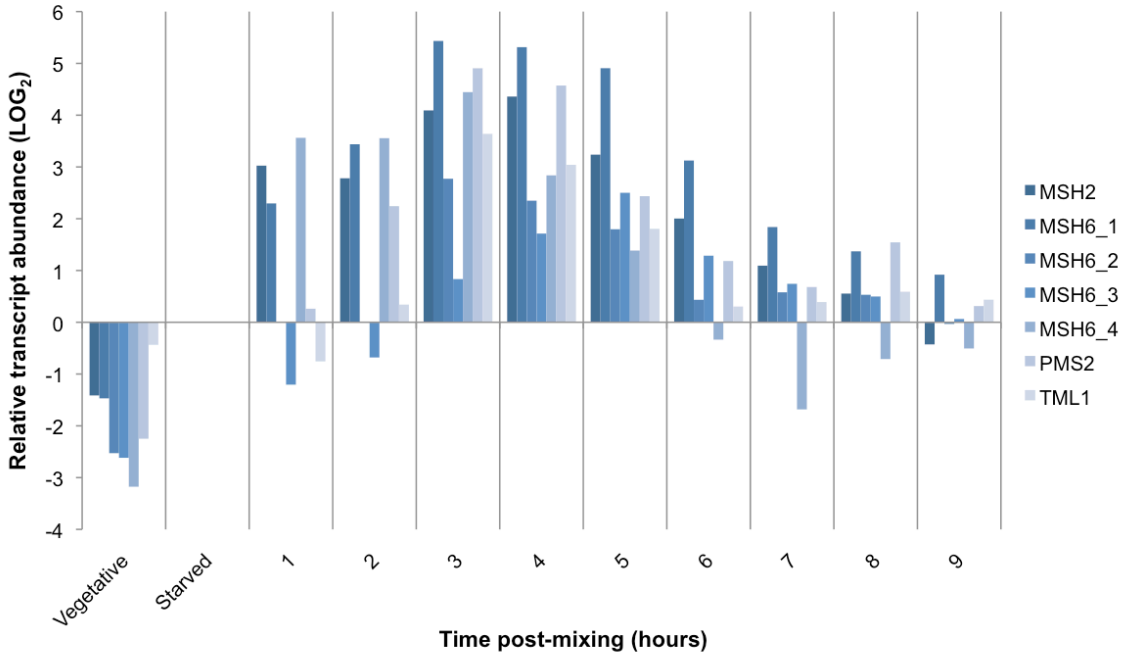


Figure 13. MMR transcript abundance during vegetative and conjugal development.
mRNA was isolated from vegetative and conjugating *Tetrahymena* cultures, reverse-transcribed into cDNA and then subjected to quantitative real-time PCR, as described in Materials and Methods. For conjugating cultures, samples were taken at hourly intervals, for 9 hours, after conjugation was initiated. Transcript abundance levels are relative to starved cells and are normalized to the *17S rRNA* and *RPL21* transcript levels. The data are the average of two biological replicates.

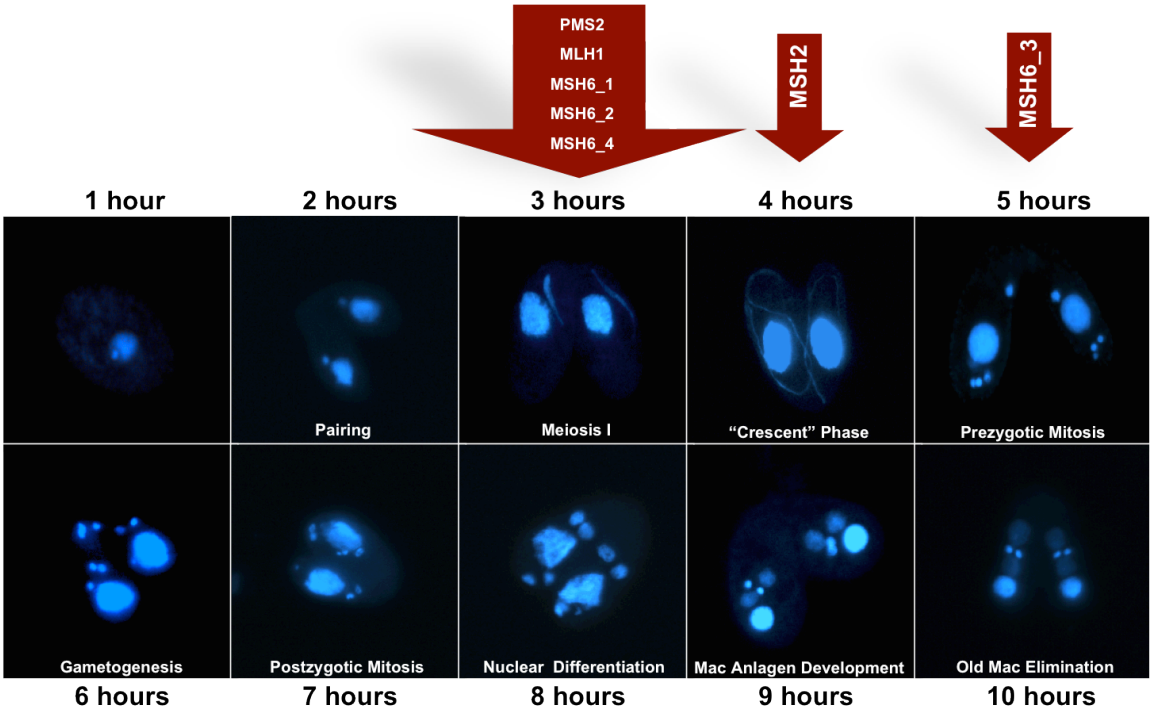


Figure 14. MMR transcript levels increase during Meiosis I, prior to nuclear exchange and gametogenesis.

Transcript levels measured throughout conjugation were determined by quantitative PCR analysis. The cytological and nuclear developmental stages corresponding to each time-point are indicated. For all MMR genes, increases in transcript abundance occur prior to gametogenesis, during meiotic prophase. Nuclear events during conjugation were visualized by DAPI staining of nuclei.

4.6 The MMR proteins are not essential for vegetative growth

Macronuclear knockouts of the *MSH2*, *MSH6_1*, *MSH6_2* and *PMS2* genes were created by disruption of each locus with the neomycin (*neo*) cassette of the p4T2-1 gene replacement vector (Figure 15) (113). This plasmid contains the histone H4-I promoter and polyadenylation site of the β-tubulin 2 (BTU2) gene, both flanking a neomycin cassette that confers resistance to paromomycin. On each side of this cassette is ~1kb of the upstream and downstream sequences that flank the MMR gene of interest. Through

homologous recombination, the MMR genes are replaced with the neomycin cassette. After biolistic transformation of both B2086.2 and CU428.2 strains with each construct (see Table 3 for genotypes), potential transformants were selected by passaging of cells every other day into media containing increasing amounts of paromomycin (150 μg/ml to 2 mg/ml) to enable complete replacement of the macronuclear MMR gene through phenotypic assortment.

During amitosis, the macronuclear subchromosomal fragments are randomly distributed to daughter cells in a process called phenotypic assortment (83). The acentromeric nature of the macronuclear chromosomes means it is not possible for cells to equally partition the DNA, yet through enough divisions, cells eventually assort to purity for one allele or the other. In the case of macronuclear gene replacements, integration of the gene replacement cassette results in expression of the neomycin gene and resistance to paromomycin. Thus, even a few copies of the allele will provide a selective advantage to cells growing in paromomycin and which contain the neomycin gene. If the gene being replaced is nonessential, than it will be possible to eliminate all wild-type copies of that gene through replacement with the selectively advantageous neomycin cassette. Those genes that are essential will not assort to purity.

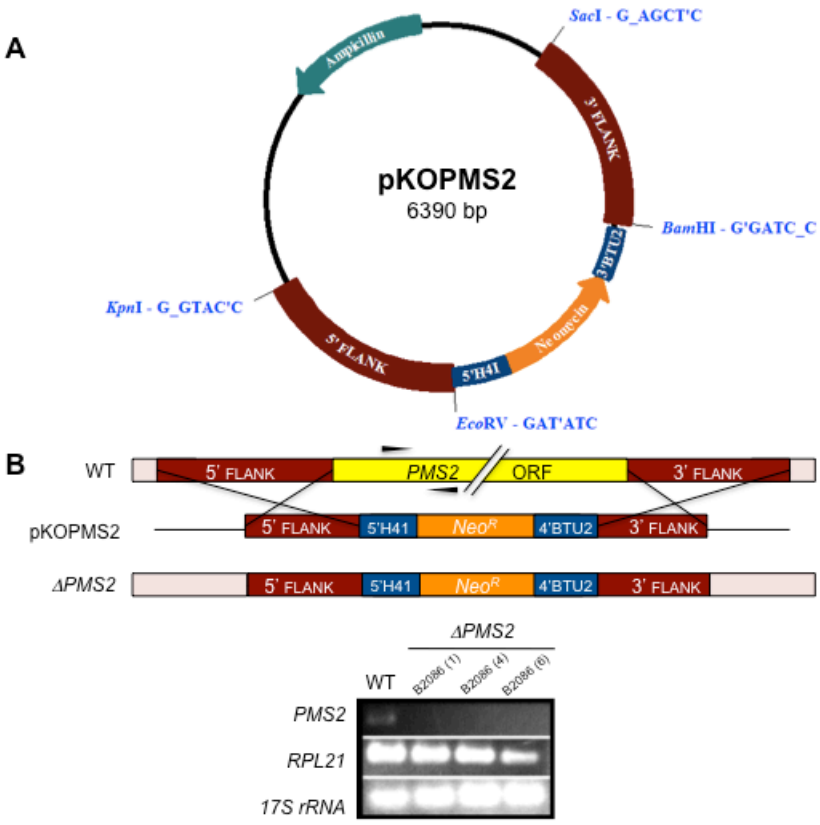


Figure 15. Macronuclear MMR gene replacement in Tetrahymena.

A. Representative map of the MMR gene replacement vector. It is constructed in the p4T2-1 vector background and contains the neomycin resistance cassette consisting of the neomycin resistance gene (Neo^R) flanked 5' by the histone H4-1 promoter and 3' by the β-tubulin 2 polyadenylation sequence. Flanking the 5' and 3' ends of the neomycin cassette are ~1 kb fragments that are homologous to those found flanking the MMR gene of interest, in this case the *PMS2* gene. Restriction sites used in cloning of the flanking regions are indicated. B. Schematic of wild-type (WT) *PMS2* locus and mutant *PMS2* locus (Δ*PMS2*) after transformation of *Tetrahymena* with the gene replacement cassette of plasmid pKOPMS2. Through homologous recombination of flanking regions, it is possible to replace wild-type alleles with the gene replacement cassette. RT-PCR analysis of mRNA transcript levels, using primers internal to the PMS2 gene, was used to confirm interruption of the coding region of *PMS2* by the gene replacement cassette. The *RPL21* and *17S rRNA* transcripts were used as controls.

Several single cell isolates were obtained for each knockout. For isolates used in this study, gene replacement of the MMR genes were confirmed by RT-PCR, as represented in Figure 15. There were no detectable transcripts of the genes of interest indicating that the gene of interest was disrupted and not being transcribed. To ensure that knockouts were complete and did not assort back to wild-type, samples of cells were grown in media without paromomycin for more than 8 months. After this time, they were challenged with paromomycin, at the high dose of 500 μg/ml, and all were found to be resistant, indicating that the neomycin cassette was still present. Wild-type cells were killed at this dose within two days, reaffirming that the resistant transformants were complete knockouts and that the *MSH2*, *MSH6_1*, *MSH6_2* and *PMS2* genes are not essential for the vegetative growth of *Tetrahymena*.

4.7 Macronuclear MMR knockouts exhibit decreased pairing efficiency during conjugation

To examine the effects of macronuclear MMR gene knockouts on the conjugation proficiency of *Tetrahymena*, cells corresponding to each knockout were grown up, starved and then mixed with a strain of a different mating type containing the same gene knockout. These matings were each performed with three separate biological isolates. Cells were sampled, fixed and counted at hourly time points and the pairing efficiency (number of cells paired / the total number of cells counted x 100%) determined (Figure 16). The Δ*MSH6_2* strains performed similarly to the wild-type B2086.2 x CU428.2 mating, with pairing efficiency quickly increasing to >80% at 4 hours, stabilizing at 5 to 7 hours, and decreasing throughout the later stages of conjugation. Interestingly, while

the initial half of conjugation shows little variability between biological replicates of Δ*MSH6_2*, the later half shows increased variability in pairing efficiency. The wild-type matings show little variability between replicates throughout conjugation. Individual mating profiles for each of the crosses are in the Appendix Figure 2.

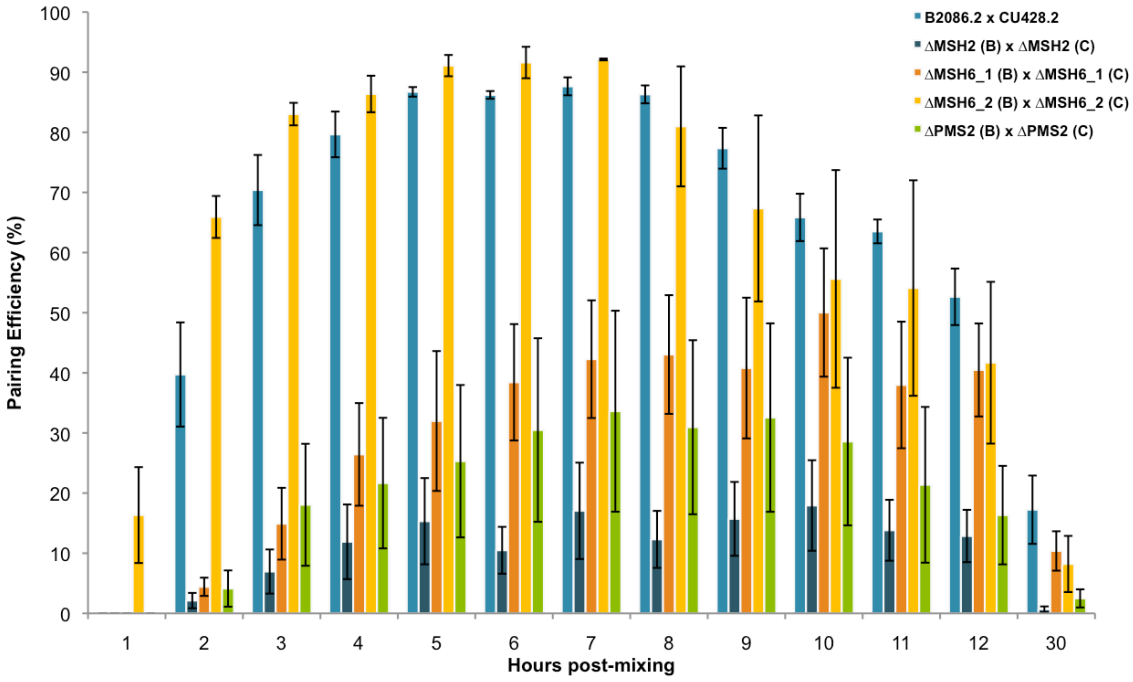


Figure 16. Pairing efficiencies of wild-type and macronuclear MMR gene knockouts during conjugation.

MMR knockout strains of complementary types were starved and mixed to initiate conjugation. Samples were taken hourly and the number of paired cells vs. total cells counted microscopically. The results shown are from three biological replicates. Error bars represent the standard error of the mean.

The profiles generated by the Δ*MSH2*, Δ*MSH6_1* and Δ*PMS2* strains contrast quite remarkably to the wild-type and Δ*MSH6_2* strains. The Δ*MSH2* strains exhibited very low pairing efficiencies, which remained relatively constant throughout conjugation and

never achieved average efficiencies of higher than 30%. The range of pairing efficiencies was varied considerably among biological replicates of the Δ*MSH2* knockout matings, as shown by the large error bars. For instance cells in the Δ*MSH2b6* and Δ*MSH2c3* cross did not pair (Appendix G), even though cells underwent all manipulations and treatments in parallel with the other Δ*MSH2* and wild-type strains, thus suggesting biological rather than technical variability. Furthermore, this cross, in preliminary matings performed six months earlier, showed delayed pairing but achieved higher pairing efficiencies (>50%), suggesting an accelerated degeneration in the Δ*MSH2* knockouts. The Δ*MSH6_1* and Δ*PMS2* knockouts had similar profiles to each other, with a delayed onset of pairing and maxima at 7 and 10 hours, respectively. The pairing efficiencies of the Δ*PMS2* were slightly lower than those of the Δ*MSH6_1*, but both similarly contained substantial variation in efficiencies, throughout conjugation. As was the case with Δ*MSH2* knockouts, the cross of Δ*PMS2b1* and Δ*PMS2c1* (Appendix G) was tested in preliminary matings four months earlier, but unlike the Δ*MSH2* cross, the pairing efficiencies of the Δ*PMS2* knockouts were the same at both points. Wild-type cells retained high pairing efficiencies in both sets of matings.

Fertility, or the ability to successfully complete conjugation, was determined for all strains, regardless of their pairing efficiencies, using a drug analysis test (Table 7). For this test, cells that had completed conjugation were grown up overnight in a final concentration of 1% proteose peptone and then added to microtiter wells containing either 15 μg/ml 6-methylpurine (6-mp) or 200 μg/ml paromomycin (pm) in parallel. Those cultures in which mating partners were fertile and able to complete conjugation

and produce true exconjugants are resistant to 6-mp, while those not completing conjugation and retaining the parental macronucleus remain resistant to pm, the drug with which they were initially selected. Before conjugation, knockouts made in the CU428.2 background are homozygous in the micronucleus for the 6-mp resistance allele (Table 2). Since the micronucleus is transcriptionally inactive during vegetative growth, expression of this allele can only occur if the micronucleus firstly, successfully completes the meiotic and mitotic divisions accompanying conjugation, and secondly, successfully differentiates one of its progeny nuclei into a new macronucleus. Both the wild-type mating and Δ*MSH6_2* matings produced clones that were resistant to 6-mp, indicating those mating cultures did contain cells that had completed conjugation. A previously made Δ*MSH6_2* knockout also showed resistance and had levels comparable to wild-type, with 87.5% of cells being resistant compared to 84% in wild-type. As expected, since it isn't possible to achieve 100% pairing efficiency, there were cells that did not mate and thus retained the parental macronucleus and paromomycin resistance (Table 7). The mating of the Δ*MSH6_1b6* isolate with Δ*MSH6_1c4* isolate resulted in 6-mp resistant cells, indicating that even though pairing efficiency was lower than in wild-type, a proportion of cells were still able to successfully complete conjugation and produce true progeny. None of the Δ*MSH2* or Δ*PMS2* strains were determined to have cells that had successfully completed conjugation, at least under the conditions tested in this experiment.

Table 7. Test for the completion of mating in MMR deficient strains of *Tetrahymena.*

Strain	Mating Pair	6-mp^R (15 µg/ml)	Pm^R (200 µg/ml)	6-mp^R ⇨ Pm^R *	Interpretation
Wild-type	B2086.2 x CU428.2	●	-	-	• Cells completed conjugation
ΔMSH2	B1 x C1	-	●		• Cells did not complete conjugation
	B4 x C2	-	●		
	B6 x C3	-	●		
ΔMSH6_1	B1 x C1	-	●		• Only B6 x C4 cross contained cells that completed conjugation
	B6 x C4	●	●		
	B13 x C6	-	●		
ΔPMS2	B1 x C1	-	●		• Cells did not complete conjugation
	B4 x C4	-	●		
	B6 x C6	-	●		
ΔMSH6_2	B1 x C1	●	●	-	• Cells completing conjugation did not retain the parental MAC or *neo* cassette • Cells <u>not</u> completing conjugation retained Pm resistance
	B4 x C4	●	●	-	
	B6 x C6	●	●	-	

6-mp^R – 6-methylpurine resistance; Pm^R – paromomycin resistance; ● = growth, - = no growth; blank = not tested;
* 6-mp^R ⇨ Pm^R – 6-methylpurine resistant cells transferred to media containing (200 µg/ml) paromomycin
B – B2086.2; C – CU428.1; numbers beside B and C indicate single-cell isolate number.

To examine if the phenotype conferred by MMR deficiency could be reversed, a conjugation rescue experiment was performed in which a MMR deficient strain was mixed with a wild-type B2086.2 strain. Here, "rescue" means an increase in pairing efficiency to near wild-type levels among mating partners. The MMR deficient strains used were those that contained the lowest pairing efficiencies among replicates tested, such that any changes in pairing efficiency could more likely be readily observed. There were increases in pairing efficiencies for all of the strains other than the Δ*MSH2c3* cross, although these increases were small and none achieved wild-type levels other than the

Δ*MSH6_2c1* cross. The Δ*PMS2c1* mating did showed a small increase in pairing efficiency to ~10% throughout conjugation, in contrast to the KO x KO cross which was unable to from mating pairs. The increase of 35% at 8 hours is anomalous. Both the Δ*MSH6_1* and Δ*MSH6_2* strains showed small increases in pairing efficiency, but that of the Δ*MSH6_1* occurred throughout conjugation and that for the Δ*MSH6_2* occurred later on in conjugation. Additional replicates of this experiment would give stronger support to these results as would additional experiments using some of the other strains available. For instance, it might be beneficial to see whether mating a CU428.2 wild-type strain with mutants made in the B background perform similarly.

In the Δ*MSH6_2b1* x Δ*MSH6_2c1* KO cross there is a steady decrease in pairing efficiency from 7 hours on, but in the conjugation rescue experiment, this decrease does not occur and most cells are still paired at 12 hours. This is suggestive that there may be a developmental block occurring at the 7 hour time point in the KO cross, causing the cells to unpair. The Δ*MSH2c3* isolate had no increase in pairing efficiency or visible pair formation when crossed with the wild-type B2086.2 strain, the same as was seen in the Δ*MSH2* knockout cross. The results of these experiments suggest that conjugating partners must both be capable of forming pairs to conjugate and that the pairing phenotypes observed are not completely rescuable.

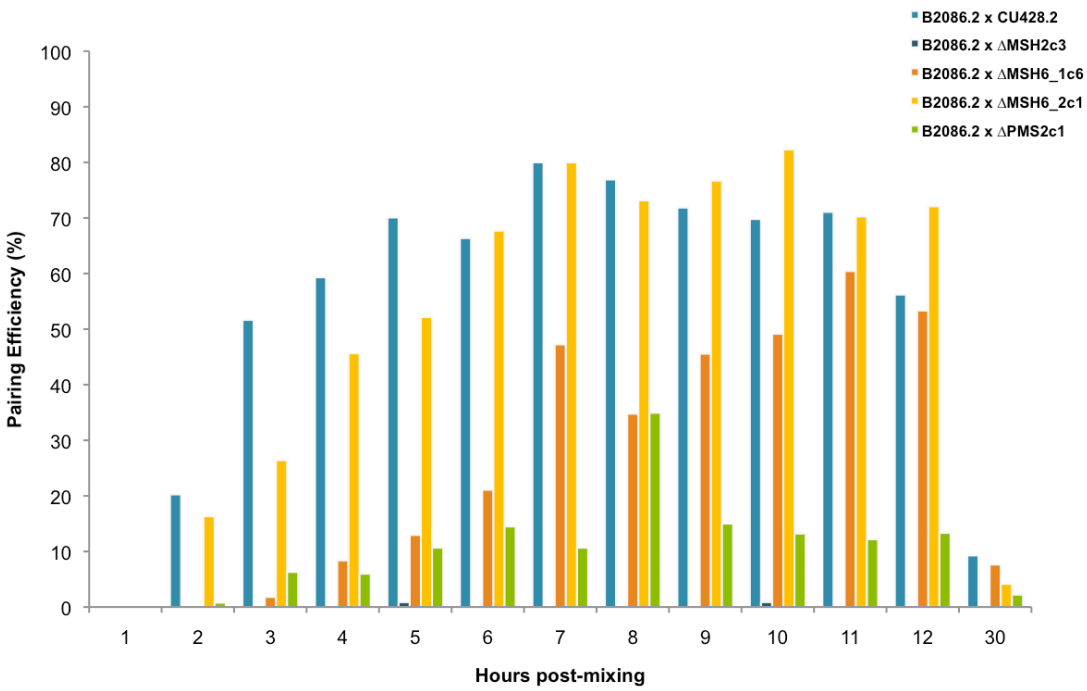


Figure 17. Conjugation rescue of the MMR knockouts in *Tetrahymena*.
The ability of a wild-type strain to rescue the pairing efficiencies conferred by knockouts of the MMR genes was tested by mating wild-type B2086.2 cells with the indicated MMR knockout strains. Hourly time points were taken throughout conjugation and the ratio of paired cells to total cells counted was determined (n=1).

4.8 *PMS2* and *TML1* knockouts lack micronuclear "target sequence"

The mutation frequency of MMR macronuclear knockouts was examined using the Random Mutation Capture (RMC) Assay (95). This assay utilizes a *Taq*αI restriction endonuclease (5'-TCGA-3') site within a locus of interest. Upon digestion of genomic DNA with the *Taq*αI enzyme, wild-type sequences are cut, while mutant sequences, in which the restriction site has been destroyed, remain intact (Figure 6). Using a qPCR

approach, single molecules harbouring mutations in this site can be amplified using primers that flank the restriction site while the wild type cannot. By determining the total number of molecules screened, using primer pairs on one side of the restriction site, it is possible to determine mutation frequency in the chosen locus. The MSE2.9 internally eliminated sequence (IES) and GRL1 gene were chosen to differentiate between MIC and MAC mutation frequencies, respectively. Since the MSE2.9 IES is only present in the MIC, it serves as an appropriate target for micronuclear mutation frequency determinations. The GRL1 target was chosen because it has been relatively well characterized and has been shown to have a negligible effect on the viability and fitness of cells when knocked out. Thus there should be no selection against mutation in this gene.

Extensive optimization of assay parameters was initially performed to ensure amplification of single mutant molecules. This included optimization of primer concentrations and melting temperatures (Table 5). PCR efficiencies for each primer pair were determined using standard curves and were >90% for all targets (Table 5). Determination of total copy number in wild-type cells indicated that in the undiluted DNA sample (5^0 dilution) there were ~5,300,000 copies of MAC target (GRL1) and ~94,000 copies of MIC target (MSE2.9), a ratio of approximately 56:1 of MAC to MIC target (data not shown). This ratio is close to the predicted 50:1 ratio of MAC alleles to MIC alleles that has been previously estimated from bulk DNA measurements, and indicates that using the purification methods employed, there was relatively little or no

loss of target, either from the MIC or MAC and the relative amounts of each were as predicted.

In order to calculate mutation frequency, one needs to know, along with the number of mutant molecules detected, the total number of molecules screened. This was accomplished in several steps. Firstly, the limiting dilution of the initial 5-fold dilution series discussed above was determined. The limiting dilution is the last dilution at which amplification is observed. For the GRL1 amplicon (MAC target), the limiting dilution was determined to be the 5^{10} dilution and for the MSE2.9 amplicon (MIC target) it was the 5^{7} dilution (data not shown). Using this dilution, and the next dilution in the series which did not amplify, the number of molecules can be estimated by performing PCR of these samples in at least 30 wells each of a 96 well plate. At limiting dilution, the number of detectable targets follows a Poisson Distribution, which describes the probability of an event taking place when the frequency of that event occurring is very low (145). When a unit volume of solution contains little to no target, the probability of having zero target is equal to the negative rate, or the ratio of the number of negative results (PCR wells that don't contain product) to the number of total tests (PCR wells screened). This is related to the copy number C, by the following equation:

$$C = -\ln\left(\frac{N_0}{N_T}\right)$$

where N_0 is the number of negative results, and N_T is the total test number. In the RMC, we are attempting to detect rare single molecules of target. Therefore in a 30-well

experiment, by determining the number of wells that don't amplify, it is possible to determine the average copy number of interest. For the GRL1 control target at a dilution of 5^10, this was ~ 0.54 copies per well and for the MSE2.9 control target at a 5^7 dilution, this was determined to be ~ 1.20 copies per well. By applying these numbers to the original dilution series, it was then possible to utilize the regression equation of the standard curve for estimation of the total target number in *Taq*αI digested DNA samples (Figure 18).

When estimating the total number of micronuclear targets for an undiluted sample, the wild-type B2086.2, Δ*MSH2* and Δ*MSH6_1* genomic DNA samples had similar copy numbers of micronuclear target, whereas the Δ*PMS2* and Δ*TML1* genomic DNA samples indicated very little and almost no micronuclear target. But interestingly, when comparing total copy numbers of the macronuclear target (Figure 18) at a dilution of 5^4, all samples contained relatively the same amounts of target. In fact, for the Δ*TML1*, the average copy number is nearly the same as wild-type levels. Thus while it appears as though the total DNA content in the samples is the same, there is a significant loss of micronuclear target and this cannot be explain by DNA degredation, since the macronuclear target copy number is relatively unchanged among all samples. We were curious as to whether this loss in target was a result of a lack of micronuclei or micronuclear DNA. Using confocal fluorescence microscopy and DAPI staining of nuclei, all cells, including the Δ*PMS2* and Δ*TML1*, were found to contain micronuclei with no decrease in micronuclear DAPI fluorescence among cells, suggesting loss of the specific MSE2.9 sequence.

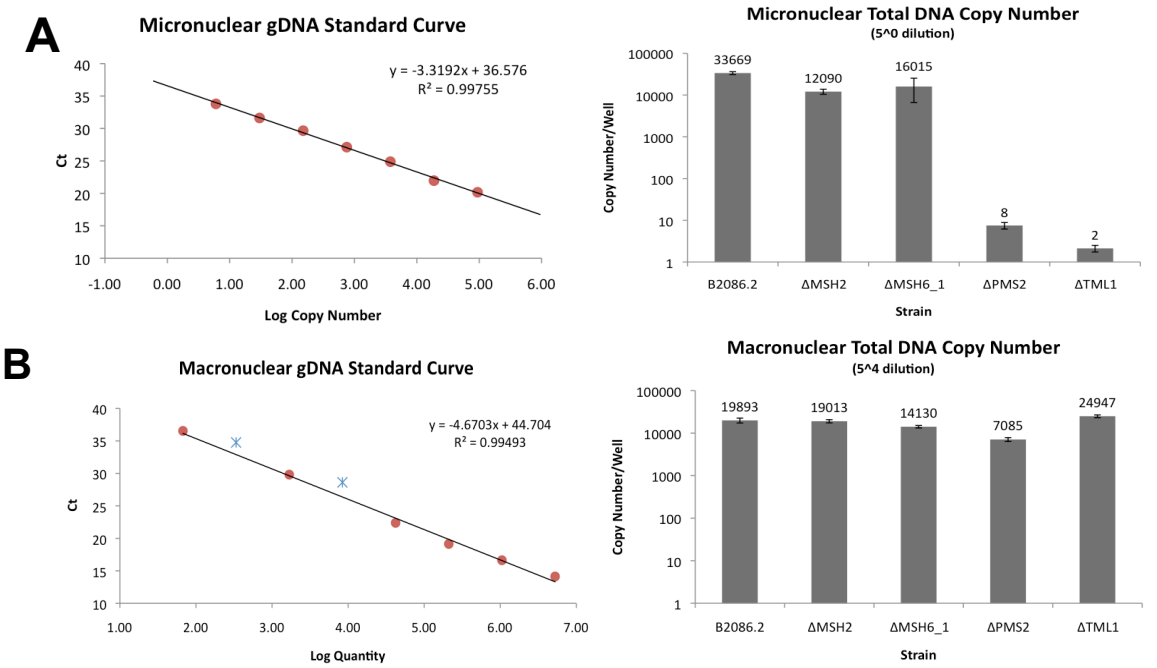


Figure 18. Macronuclear and micronuclear total DNA copy number in the RMC assay. Using genomic DNA standard curves and the zero order equation of the Poisson distribution, the average copy number of (A) micronuclear and (B) macronuclear targets was determined at the 5^0 and 5^4 dilutions, respectively. Blue markers on the macronuclear gDNA standard curve indicate points that were excluded from analysis as they were not in the linear range. Error bars represent the standard deviation of triplicate measurements.

Wild-type and MMR knockout samples were screened for molecules containing mutations in the *Taq*αI site using primers that flanked the restriction endonuclease site. We were able to detect product amplification in wells at the appropriate, limiting dilutions for micronuclear and macronuclear targets. However, all of the amplicons generated for the MSE2.9 and GRL1 targets cleaved with *Taq*αI. Table 8 indicates the number of putative mutant molecules detected in each dilution tested, before digestion with *Taq*αI. For the micronuclear targets (MSE2.9), all samples were run on agarose gels to determine putative mutant numbers. For the macronuclear target (GRL1), numbers were determined based on melt curve analysis and those sample run on agarose gels. All

putative products were cleaved by *Taq*αI indicating that these were either false positives or mutants that had regenerated the restriction site during PCR amplification.

Table 8. Putative numbers of micronuclear and macronuclear mutant molecules in dilutions of *Taq*αI digested genomic DNA from various strains of *Tetrahymena* using the RMC assay.

	gDNA DILUTION and TARGET					
	Micronucleus (MSE2.9)			Macronucleus (GRL1)		
Strain	5^0	5^1	5^2	5^3	5^4	5^5
B2086.2	--	62	0	--	11	1
SSMSH2b1	--	0	--	>48	15	--
LSTML1b	17	--	--	30	32	2
SSPMS2b1	0	--	--	--	11	8

Melt curve analysis is performed after qPCR and is used to follow the dissociation and hybridization of double-stranded DNA as the temperature changes (146). This is detected using intercalating DNA binding dyes that fluoresce maximally when bound to dsDNA. Thus when DNA melts, the dye no longer binds and this is detected as decrease in fluorescence. Upon hybridization of the DNA strands, the dye is able to once again bind the DNA and this is detected as an increase in fluorescence. Differences in DNA sequence among amplicons results in different melting profiles for each thus facilitating the identification of specific amplicons during PCR. All products in the RMC assay contained a single melt curve peak except for the 518 bp amplicon (Figure 19). This amplicon was generated using the GRL1*Taq*I primers that flanked the *Taq*αI site in the GRL1 locus and it produced a bimodal melting curve. This bimodal curve, although

somewhat unusual in melt curve analysis, is most likely a result of the asymmetric GC distribution in the amplicon (Figure 20, (147)). This was also supported by analysis of the GRL1 amplicon using the "melting curve" simulation at stitchprofiles.uio.no, which predicted a bimodal melting profile (Figure 21, (148). Both experimental and simulated profiles show two transition points, corresponding to two regions of the amplicon melting at different temperatures due to localized differences in AT/GC content; AT rich regions melt before GC rich regions.

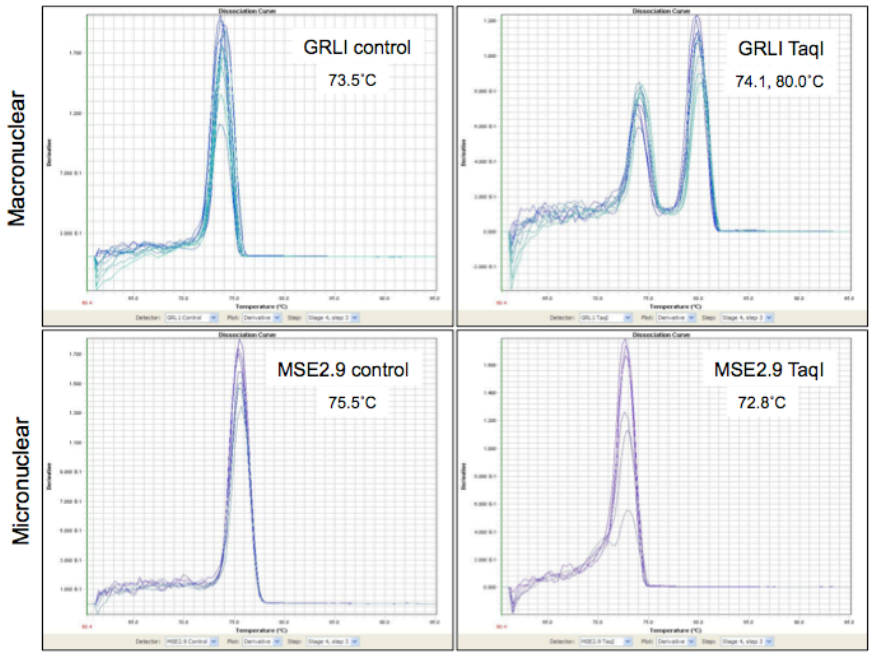


Figure 19. Melt curve analysis of amplicons generated in the RMC assay.
Dissociation curves of the macronuclear (GRL1) and micronuclear (MSE2.9) control and *Taq*αI-site containing amplicons were generated using STYO9 detection in a real-time qPCR experiment. Targets are indicated as well as their corresponding melting temperatures used to distinguish desired products from primer dimers and non-specific amplicons generated at low dilutions of template. Control targets refer to those amplicons generated with primers upstream of the *Taq*αI site and *Taq*I targets refer to those amplicons generated with primers flanking the *Taq*αI site.

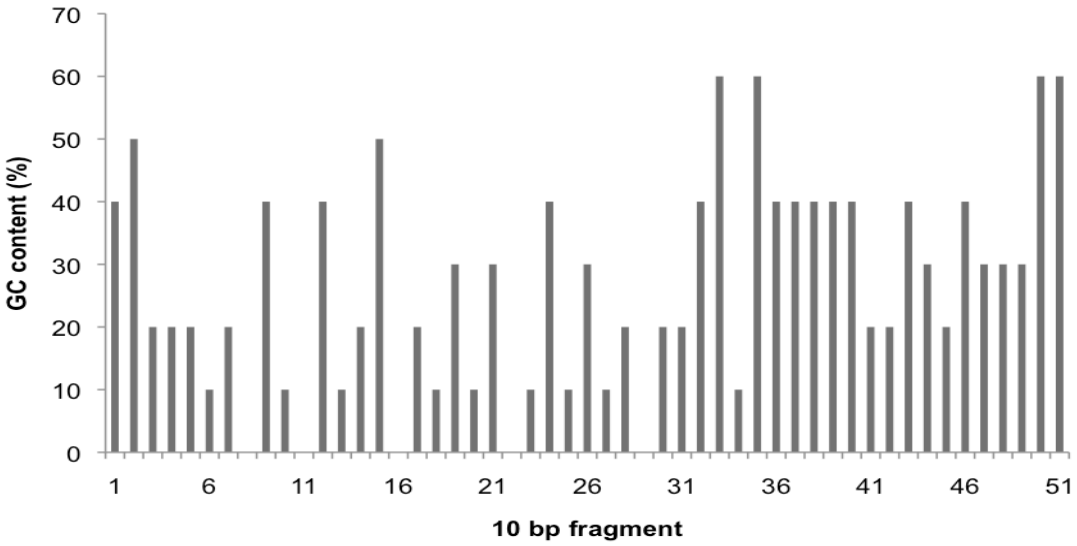


Figure 20. GC distribution along the GRL1 *Taq*αI amplicon template sequence.

A comparison of GC content is shown per 10 base pairs along the template sequence of the 518 base pair GRL1 amplicon containing the *Taq*aI restriction endonuclease site. The GC content is lower in the initial thirty 10 bp fragments (300 bp) than it is in the latter half of the molecule.

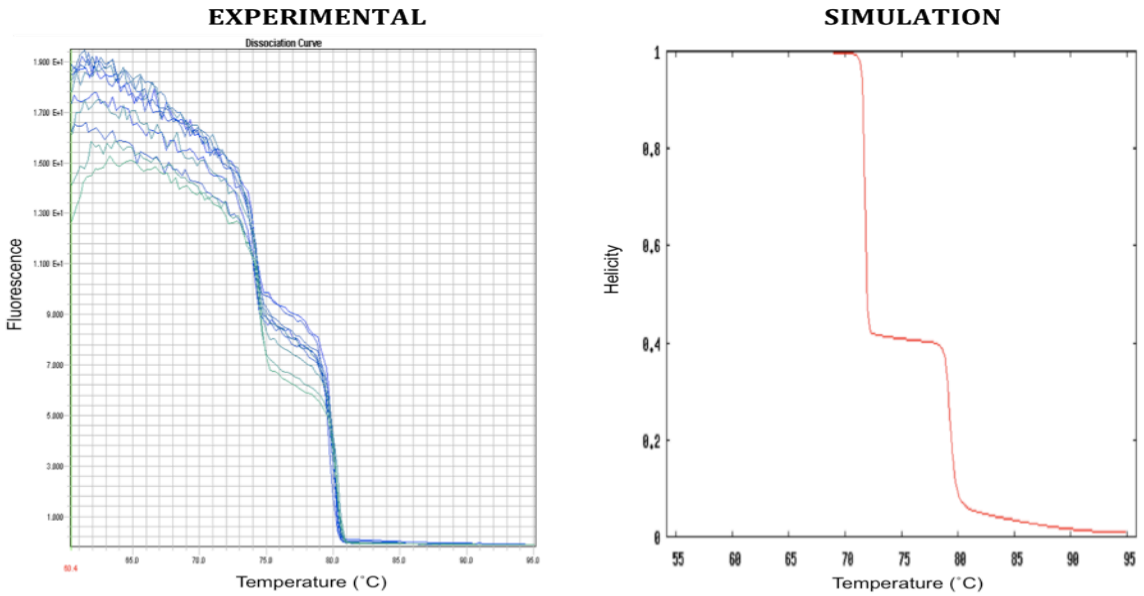


Figure 21. Experimental and simulated melting curves for the GRL1 *Taq*αI amplicon.

Melting curve profiles for the 518 bp GRL1 *Taq*αI amplicon show a bimodal transition in experimentally determined plots and a simulated plot obtained using stitchprofiles.uio.no (148).

Chapter 5: Discussion

5.1 *MSH6* expansion in *Tetrahymena*: a ciliate specific occurrence?

The primary objective, to identify putative MMR homologues in *Tetrahymena,* was fulfilled with the identification of two MutL homologues, PMS2 and TML1, and five MutS homologues (an MSH2 and an atypical four MSH6 homologues). As is characteristic of all MutS homologues, the highest sequence conservation of the *Tetrahymena* MSH6 homologues was within the ATPase domain. Further analysis of the protein sequences and structures indicates that while the functional domains of the proteins are generally conserved, only the MSH6_1 protein contains the PCNA-binding motif characteristic of MSH6s in other organisms (128). This suggests that the MSH6_1 homologue is functional in the post-replicative repair of DNA since interactions with PCNA are important for targeting the MSH6 protein to replication foci and activating the endonuclease activity of MutLα (128, 63). Interestingly, we have also observed that while the MSH6_1 protein localizes to the micronucleus during vegetative development and conjugation, the MSH6_2, MSH6_3 and MSH6_4 do not, suggesting that divergence of the MSH6 paralogues has resulted in functional differences during the cell cycle and conjugation. In *Tetrahymena*, the importance of MSH6_1 is further supported by the observation that macronuclear knockouts show a decrease in pairing efficiency during conjugation.

Tetrahymena isn't unusual in containing multiple MSH6 homologues, but is rather unusual in the number of them. *Arabidopsis thaliana, Solanum lycopersicum* (tomato)

and *Zea mays* all contain a second MSH6 paralogue (related by gene duplication events), called MSH7 (mus2 in *Z. mays*), but this orthologue (related by speciation events) hasn't been identified in any other eukaryotes to date and is likely plant specific (149-151). The finding that other ciliate species, including *Ichthyophthirius multifiliis*, *Oxytricha trifallax* and *Paramecium tetraurelia* all include multiple MSH6 homologues is unprecedented. We speculate that ciliates have evolved an expanded set of MSH6 homologues in order to increase functionality in other processes. This expansion may be reflective of the unique genetics of ciliates, which possess two genomes, one in the transcriptionally-silent germline micronucleus and the other in transcriptionally-active somatic macronucleus. While the micronucleus divides by mitosis and meiosis, the macronucleus divides by amitosis, with acentromeric chromosomes partitioning via phenotypic assortment (152). In addition, during conjugation, the new macronucleus undergoes extensive genome rearrangements and fragmentation while the old parental macronucleus degrades through an apoptotic-like process termed programmed nuclear death (91). The unusual complement of MSH6 proteins in *Tetrahymena* and other ciliates suggests that each homologue, in concert with MSH2, may have a specialized role in one or more of these biological phenomena.

Expansion of *Tetrahymena* gene families is not unique to MSH6 and is unlikely to be due to whole genome duplication. In an analysis of the macronuclear genome sequence, Eisen and colleagues report that although expansion of gene families is abundant in *Tetrahymena*, the duplication events associated with these expansions occurred at distant times, with some being ancient and some being more recent (80). Furthermore, Xiong

and colleagues identified 165 genes corresponding to putative ABC transporters and further analysis indicated that gene duplication and sequence divergence has led to a highly expanded set of eight ABC transporter families (153). Of particular interest, is the fact that these transporters, as well as all known MutS proteins, are members of the ABC-ATPase superfamily (154). This family of proteins is extensive and includes members that utilize ATP hydrolysis to provide energy for the transportation of nutrients across membranes, the detection of toxic chemicals and pathogens and the repair of DNA. Thus expansion of component family members, including the MutS homologues, may be a means by which *Tetrahymena* has been able to adapt to its environment. Interestingly, it has also been demonstrated that protein evolution in ciliates is driven by genome architecture, specifically chromosomal processing (155). Those ciliates with extensive chromosomal processing and highly fragmented genomes have an increased rate of protein evolution and numbers of divergent paralogues. Although the extent of chromosomal processing in *Tetrahymena* is less than in other ciliates, this, along with the previous finding, may be one of several factors contributing to the number of MSH6's present in *Tetrahymena*.

5.2 Measuring mutation frequency in *Tetrahymena*: Not all it's "cut" out to be.

The MMR repair system elevates the fidelity of DNA replication between 50 – 1000 fold (32). Thus, a primary consequence of cells in which the MMR genes have been either deleted or functionally inactivated is an increase in mutation. Our objectives following the identification the putative MMR homologues in *Tetrahymena,* were to

firstly create macronuclear knockouts of the MMR genes and secondly to determine whether cells in which the MAC copies of the MMR genes were disrupted had an increase in mutation. Specifically, we were interested in investigating potential differences in mutation frequencies in the germline MIC genome versus somatic MAC genome, The measurement of mutation can be expressed in two ways, either as a mutation rate or a mutation frequency. Mutation rate is expressed as the number of mutants arising per generation and can be determined either through mutation accumulation experiments or fluctuation analysis (156). Mutation frequency is expressed as the number of mutants in a population and can be estimated using fluctuation analysis or assays such as the LacZ reversion assay in *E. coli* (92, 93), or Big Blue mouse assay in mice (157). While these assays are well established, they all require the observation of a phenotype associated with the mutational target. The study of mutation frequency is significantly more challenging in *Tetrahymena* for a number of reasons. Firstly, the use of *Tetrahymena* in phenotypic assays is not possible due to a lack of available macronuclear markers. Secondly, the absence of transcription in the MIC during vegetative growth hides any phenotype provided by the micronucleus. All of these properties necessitate the use of a genetic assay like the RMC rather than a phenotypic assay. Macronuclear mutation accumulation has been studied in *Tetrahymena* by following extinction of clonal cell lines using a competition assay with $CdCl_2$-inducible GFP strains of *Tetrahymena* (158), but, as with the other assays, this can only measure somatic mutation rate and not germline mutation due to the absence of phenotype contributed by the MIC. Additionally, cadmium (Cd^{2+}) is a potent inhibitor of mismatch repair (159, 160), therefore using the MA assay as described would confound any results.

In order to provide another means by which to measure mutation frequency in a population of cells, Jason Bielas and Lawrence Loeb developed the random mutation capture assay (RMC) in 2005 (94). This assay detects and quantifies spontaneous random mutations occurring at an enzyme-specific site in a chosen gene, thus allowing for the determination of mutation frequency. The principal advantage of using this assay over those described above is that it is a genetic assay and not a phenotypic assay, thus any locus within the genome can be studied.

Since its publication, there have been numerous reports of the RMC being employed, but the majority of these reports have been performed using mitochondrial DNA and only a few use genomic DNA (for an expansive list, see Supplementary Table 2 in (161)). The RMC assay potentially overcomes the challenge posed by nuclear dimorphism of *Tetrahymena:* the separation of mutation events occurring in MIC from those occurring in the MAC. Although the MAC genome is derived from that of the of MIC, the MIC contains multiple internally eliminated sequences (IESs) that are absent in the MAC. Through an RNAi-based mechanism, thousands of these sequence are eliminated from the MAC during macronuclear development (162); we used the 2.9 kb MSE2.9, a well characterized IES (96, 163, 164, 89), to detect mutation in the MIC. Conversely, because the MAC is derived from the MIC, there are no MAC specific sequences to utilize as targets for measuring mutation frequency. While it is possible to generate nuclear preparations that are highly enriched for macronuclei, complete purification is not possible(165). Additionally, it was our goal to develop a relatively quick and simple

assay in *Tetrahymena* that would not require such labour intensive preparations. Therefore, we chose to use the relatively well characterized GRL1 gene as a macronuclear marker (97), assuming that the ~50-fold amplification of MAC genes would result in most of the signal coming from MAC DNA.

Extensive optimization of the RMC protocol was absolutely necessary in our hands, primarily because we were using two different targets in nuclei consisting of entirely different ploidy and DNA content. For example, the very low AT richness of the MSE2.9 region necessitated the use of lower melting temperatures during PCR than the more GC rich GRL1 amplicon. The initial attempt at the RMC by me and my colleague, Yaroslava Polosina, provided some significant challenges. We initially used the 2005 protocol (94) which relied on using DNA probes to isolate the target DNA molecules of interest from the bulk gDNA through hybridization. These probes were constructed via PCR of the region of interest, using a mixture of dUTP and dTTP. Upon treatment of samples and PCR products with uracil DNA glycosylase (UDG), which cleaves the uracil base from the uracil-containing DNA, it is possible to remove the probe and contaminating PCR products that have been amplified with incorporation of dUTP. In our experience, use of dUTP was not successful. It's been shown that incorporation of dUTP into DNA probes inhibits hybridization (166). This may also help explain the low yield of target that was purified (Yaroslava Polosina, personal communication).

As a means of pre-testing experimental conditions and determining whether the assay would be sensitive enough in our hands, we decided to apply the RMC assay to

determining mutation frequency in *E. coli.* This seemed like an ideal choice as a "control" experiment since mutation in repair-deficient strains of *E. coli* is well characterized (92, 93) and genetically, it is a much simpler system than *Tetrahymena*. Unfortunately, and to our surprise, this proved impossible to do in *E. coli*, because of contamination of the *Taq* polymerase by bacterial DNA. Probably one of the most overlooked properties of commercially purchased recombinant enzymes is the fact that they've usually been purified from *E. coli*. As such, they contain large amounts of contaminating *E. coli* DNA. *Taq* polymerases and other PCR reagents are no exception (167, 168). There are commercially available polymerases that claim to be DNA-free, but the problem of DNA contamination continued to be a problem in our hands. DEAE-cellulose treatment (169) of the polymerase was somewhat useful in removing the DNA contamination, but resulted in dilute enzyme with decreased activity compared to untreated samples (personal observation). This method of purification was impractical for use in the RMC assay because optimal polymerase activity is required during PCR in order to detect single DNA molecules. Furthermore, any restriction endonucleases, such as *Taq*αI are also recombinant, thus would require purification as well.

Using the simplified RMC assay, developed in 2011 (95), we were able to obtain amplification of all targets in all of the *Tetrahymena* strains tested, although the micronuclear MSE2.9 target in the Δ*PMS2* and Δ*TML1* strains was very low due to apparent loss of micronuclear target (see below). Surprisingly, all of the amplicons (putative mutants) detected in the assay appeared to be false positives, meaning they were cut by the *Taq*αI restriction enzyme, a property of wild-type sequences. There are

multiple explanations as to why this may have happened. The first obvious explanation is incomplete digestion of the gDNA. Although possible, the amounts of enzyme and digestion times used are excessive, with a 10-fold increase in the number of units suggested by the supplier (NEB). Another potential explanation is the regeneration of wild-type sequences from mutant templates. The mutational targets are swamped out by the large number of wild-type sequence fragments (i.e. that have been digested by *Taq*αI) and thus we hypothesize that these fragments may act as primers in subsequent reactions. It is known that in PCR, reannealing of template strands can outcompete binding of primer thus affecting the dynamics of the PCR (170). In a sample that is highly enriched for a target template (i.e. macronuclear GRL1 gene target), it seems likely that this could happen at an early stage of the PCR, preventing binding of primer and possibly regeneration of the WT sequence. Based on these results, it would be beneficial to isolate the target amplicon from the bulk DNA by optimizing and adapting the probe synthesis and hybridization for the current protocol. It would perhaps also make it easier to detect mutation in the micronucleus, which like the mutant molecules in the MAC genome, are fewer in number relative to the MAC fragments in the sample, a result of the MIC being diploid and the MAC being polyploid.

A third potential factor contributing to the lack of mutant detection is the use of the *Taq*αI site, which consists of only four bases (TCGA). To more accurately determine mutation, one would require screening multiple restriction sites and multiple genes for each sample, since mutations occurring outside of this restriction site are not detected in the assay, resulting in underestimates of mutation frequency (171, 172).

The RMC assay, while potentially underestimating mutation frequencies, has proven to be informative in other organisms, particularly in describing mutation in cancer and aging (39, 173). Although it has been challenging to establish the RMC assay in *Tetrahymena* this is not wholly a consequence of technical challenges, but rather may be a consequence of the unique and complicated biology of *Tetrahymena*.

5.3 Potential roles of the *Tetrahymena* MMR homologues during vegetative and sexual development

Initial attempts at creating MMR knockouts of the *MSH2, MSH6_1*, *MSH6_2* and *PMS2* genes, while successful in terms of MMR gene replacement, surprisingly resulted in strains that were infertile and/or unable to form mating proficient pairs within a relatively short period of time after gene deletion. Initially, it was thought this was due to the natural progression of the cells, since it is known that over time with successive cellular divisions, T*etrahymena* will become infertile. Infertile strains are unable to pass their parental micronuclear DNA and genotype onto progeny cells. This is tested by challenging cells with either 6-methylpurine or cycloheximide. Strains carrying this drug resistance marker are homozygous for it in the MIC. Only progeny cells that obtain this marker during conjugation will become resistant, since the phenotype of the MAC, and thus the cell, are derived from the MIC. The inability of cells to successfully complete mating is most likely a result of mutation accumulation and micronuclear degeneration (174). Since initial knockouts were unable to conjugate, they were created again in the same background wild-type strains. To decrease the number of cellular divisions between

experiments, all strains were kept in soybean culture and overlaid with mineral oil (116). Interestingly, limiting cellular divisions did not prevent MAC KO strains from becoming infertile or unable to form mating proficient pairs. The only knockout strains that appeared unaffected were the Δ*MSH6_2* isolates, and yet they were the first KO's to be constructed, along with the Δ*PMS2* strains, which, if able to pair, showed considerably delayed and decreased pairing efficiencies throughout conjugation. The Δ*MSH2* and Δ*MSH6_1* strains were also both prepared at the same time, months after the previous two. The oldest of the strains were the wild-type strains, from which all the knockouts were made, yet they still remained capable of pairing and completing conjugation. Thus, it seems unlikely that the cause of infertility and lack of pairing seen among the MMR KO's is a result of the natural ageing process, but rather an accelerated ageing through mutation accumulation. This is further supported by the finding that MMR KO's don't perform the same in terms of pairing efficiency during conjugation and all show increased variability among biological replicates. This suggests that the observed plasticity of the MMR-deficient strains is not an immediate result of MMR deficiency, but rather a time-dependant result; most likely due to the accumulation of random mutations over time. Furthermore, the lack of certain MMR proteins prior to conjugation may affect certain protein-protein interactions that are needed for downstream processes including cell pairing and mating competence.

The likelihood for the *Tetrahymena* MMR proteins being actively involved in a variety of cellular processes is almost certain, based on what we know from studies in other organisms. Exclusive participation in the post-replicative repair of DNA mismatches is

overshadowed by the extensive interactions these proteins have with other DNA repair systems and mutation avoidance pathways (7, 33). The experimental results presented in this thesis provide partial insight into the roles these proteins may have throughout the life cycle of *Tetrahymena*.

During conjugation, gene expression in the new macronuclei (anlagen) begins around the same time it ceases in the old macronucleus, soon after the second postzygotic division at ~7-8 hours (Figure 5, Figure 14) (175). Likewise, protein synthesis decreases during postzygotic development and increases around ~7-8 hours at the first appearance of anlagen (142). Thus, based on analysis of the MMR macronuclear knockout phenotypes, inferences on the roles of the MMR proteins during conjugation can only be made prior to nuclear differentiation, although, there is support, based on the qPCR studies described and GFP-localization studies (performed by Erin J. Annandale) that they may play a role in these later processes. The following sections will describe the potential processes in which the *Tetrahymena* MMR proteins may or may not function and why. It will begin with vegetative growth and the transition to conjugation and follow with meiotic recombination and macronuclear differentiation events occurring during conjugation.

5.3.1 MMR in vegetative growth and the transition to conjugation

One of our primary goals was to discern the roles of the MMR proteins in the two nuclei of *Tetrahymena*. Our hypothesis, that the unique properties of the macronucleus obviate the requirement for MMR repair, is based on several observations. The macronucleus is a polyploid nucleus, containing ~250 subchromosomal fragments each

maintained at an average number of ~45 copies per MAC (80). Through phenotypic assortment, alleles randomly assort to daughter cells, eventually becoming homozygous for a particular allele at all loci. A mutation in a single copy of an allele, or even a few for that matter, would cause little if any harm to the cell, as >40 wild-type sequences still remain in the macronucleus. If the mutation is deleterious, it can eventually be removed from the population through phenotypic assortment. Thus MMR would seemingly be unnecessary for mutation avoidance in the macronucleus. This is supported by both qPCR studies and GFP-localization studies of the MMR proteins.

The vegetative levels of transcript, relative to starved cells, are low for all the MMR genes, indicating that during vegetative growth, only a basal level of transcription is occurring. Starved cells go through a rapid decrease in transcription within the first two hours of starvation (142), then remain at a basal level until the mixing of cells of complementary mating type and the start of conjugation. The low levels during starvation are what would be expected in cells that are not actively replicating their DNA or dividing. But in vegetative cells, since DNA replication is occurring in both the MIC and MAC and cells are actively dividing, this would be expected to correspond with a higher amount of transcription overall. Localization studies of GFP-tagged MMR proteins indicate that the TML1 protein localize only to the macronucleus during vegetative growth, which, although unlikely, could be acting as a homodimer in the MAC to function in mismatch repair. In other eukaryotic organisms the TML1 homologue interacts with PMS2 to form the MutLα heterodimer. It is also possible that the C-terminal fluorescent tag of TML1 interferes with MutLα function (176) and prevents

correct localization to the MIC. The MSH2, MSH6_1 and PMS2 proteins localize exclusively to the micronucleus during vegetative growth. Like the TML1, it may be possible that the PMS2 is functioning as a homodimer in the micronucleus, but again this is unlikely since PMS2 and TML1 interact to form a heterodimer in other organisms. The coincident localization of the MSH2 and MSH6_1 to the micronucleus is suggestive of a heterodimeric interaction between the two. The absence of any MutS homologue localization to the MAC coupled with the low transcript abundance during vegetative growth suggests that there is no MMR in the MAC. In contrast, the MIC is the germline nucleus, from which all progeny nuclei (synkaryons) are derived. In the absence of gene expression, there is no selection against nuclei containing deleterious mutations. Thus MMR is undoubtedly more important in this nucleus, helping to maintain it's genetic integrity so that progeny are viable and able to propagate.

Further evidence of the MMR proteins having an important role in micronuclear genomic integrity comes from measuring total copy number of the MSE2.9 target in the RMC assay. The Δ*MSH2* and Δ*MSH6_1* strains produced total target numbers that were in the same magnitude as that obtained for the wild-type B2086.2 strains, yet the Δ*PMS2* and Δ*TML1* strains contained very little and almost no detectable target, respectively. This was confirmed not to be a degradation issue, as macronuclear targets were present in abundant amounts and were similar to wild-type levels. Thus it appears to be a target and strain specific feature, i.e. the inability of primers to find and anneal to their target, possibly due to genome rearrangement occurring in the micronucleus of *PMS2* and *TML1* macronuclear knockouts. It would be beneficial to use primers to other IES targets within

the micronucleus to determine whether the lack of target we see is unique to the MSE2.9 IES or common to all IESs tested. Similar results among IES targets would suggest a role of the PMS2 and TML1 proteins in the maintenance of IES stability within the MIC. In yeast, the MMR proteins play an important role in preventing the recombination of divergent sequences and hence help to prevent genome rearrangements from occurring (177). The possibility that strains were losing micronuclear DNA was ruled out through microscopic analysis of DAPI stained nuclei, which revealed that cells contained brightly staining micronuclei, similar as to those in wild-type cells.

The pairing efficiency profiles of the Δ*MSH2*, Δ*MSH6_1* Δ*PMS2* strains were all similar to each other in that they were very low when beginning conjugation and increased slightly during conjugation, but never reached wild-type levels of pairing. Additionally one each of the MSH2 and PMS2 knockout strains was unable to pair when initiated to start conjugation, although partial rescue of the PMS2 phenotype was achieved when mated with a wild-type strain. Infertile strains are unable to successfully complete conjugation, meaning they do not successfully transfer parental MIC DNA to progeny, but they keep that capacity to form pairs. Thus the decline in pairing efficiencies and eventual abolishment of pairing ability is suggestive of formation of a developmental block prior to pairing, and may result from mutation accumulation of degeneracy of the MIC, the germline nucleus.

At the beginning of conjugation, the micronucleus is found in G2 (4C) and the macronucleus in G1 (~45C). Thus during starvation, the micronucleus replicates it's

DNA one more time in preparation for conjugation, while the macronucleus maintains a G1 copy number. Cells not in macronuclear G1 upon the initiation of conjugation are unable to pair and must undergo one more division before pairing can take place. It is hypothesized that the phenotype we observe, i.e. the decline in pairing efficiency to a state of pairing inability, is a result of cells either unable to enter macronuclear G1 or most likely unable to enter micronuclear G2. It seems odd that a cell that undergoes so many events during conjugation must replicate it's micronuclear DNA hours before conjugation even begins. Why is that DNA replication of the MIC can't occur upon pairing of cells? This leads to the possibility of this early replication acting as a checkpoint in and of itself. Perhaps the commitment to replicate the micronuclear genome is a commitment to forming mating proficient pairs and ultimately a commitment to conjugate. In the event of increased mismatch, as would be the case in MMR knockouts, it could be that DNA synthesis stalls, likely activated through a DNA damage checkpoint. Although, at the same time, this is equally unlikely, as cells deficient in MMR are unable to respond to DNA damage through the DNA damage and cell cycle checkpoints. This is supported by the observation that certain types of cancer are resistant to chemotherapeutic drugs. On the other hand, and perhaps a little unorthodox, it could be that the absence of one the four canonical MMR proteins creates a break in a signalling cascade not involved in DNA repair, but perhaps in a "preparatory" checkpoint involved in preparing the cell for the multiple rounds of DNA replication, meiosis and mitosis it is about to undergo during conjugation. If this is the case, it would not be the first time an unusual checkpoint has been identified in *Tetrahymena* (178).

5.3.2 MMR and the regulation of meiotic recombination

The role of the MMR proteins in meiotic recombination has been extensively investigated (179, 180), but the complexities involved in this process leave many questions and mechanistic details unanswered. Our studies indicate that during conjugation of wild-type *Tetrahymena* cells, transcript abundance of the MMR homologues is increased, relative to starved cells, during the first half of conjugation. This increase coincides with meiotic prophase of Meiosis I and prezygotic development, in which homologous chromosomes pair and are aligned to allow for homologous recombination to take place, and a single pronucleus is selected to be exchanged with the mating partner. Inspection of microarray data indicates a similar pattern of expression for all the MMR genes (181). Since mRNA levels are not directly proportional to protein levels, it seems likely that this increase in transcript abundance may rather be necessary in later stages of conjugation, and may be more reflective of the cell stockpiling protein for use in later stages, for example during macronuclear anlagen development, when multiple rounds of DNA replication and extensive genome rearrangement are occurring.

In many eukaryotic organisms, meiotic recombination is regulated by a variety of mismatch repair proteins including MSH4, MSH5, MLH1 and MLH3. The MSH4 and MSH5 proteins have no apparent function in mismatch repair and have only been shown to participate in meiotic recombination. During meiosis, the MSH4/MSH5 heterodimer forms a sliding clamp that embraces the homologous chromosomes, possibly stabilizing the Holliday-junction intermediate and promoting crossing over (182). Although the MLH1/MLH3 heterodimer does have DNA repair activity, it is also involved in meiosis

and through its associations with MSH4/MSH5 controls crossing over and crossover interference (179). Crossover interference is the process that ensures two crossovers do not occur in neighbouring regions along the same chromosome and are equally spaced apart (183). The proteinaceous synaptonemal complex (SC) mediates this process in most organisms (184), with the MSH4/MSH5 and MLH1/MLH3 heterodimers facilitating formation of the SC and its localization at crossovers (185).

In *Tetrahymena*, meiotic recombination occurs much differently than that observed in most other organisms. One of the most significant differences is the absence of a synaptonemal complex (186, 187). The SC is responsible for synapsing, or pairing and fusing, the homologous chromosomes during meiosis. In *Tetrahymena*, the micronucleus elongates to form a thin crescent, promoting the parallel alignment of homologous chromosomes in which homologous regions are adjacent to one another. The absence of an SC in *Tetrahymena* correlates with the lack of conserved MSH4, MSH5 and MLH3 genes. Thus, while crossover interference does exist in Tetrahymena (187), the mechanism must be considerably different.

It may be possible that the MSH2, MSH6_1 and PMS2 proteins of *Tetrahymena* have a role in meiotic recombination, but based on the above information, it seems unlikely. Furthermore, localization data of GFP-tagged MSH2, MSH6_1 and PMS2 indicates the MSH2, MSH6_1 and PMS2 localize to the MIC during conjugation and are subsequently found in postzygotic nuclei destined to become MICS (Figure 22 and Figure 23) performed by Erin J. Annandale). A recent paper from Josef Loidl's lab also

demonstrates that the TML1 protein (called MLH1 in their paper) also localizes to the MIC during meiosis (188). Early conjugation is a time of significant and ongoing DNA synthesis in the micronucleus, specifically during Meiosis II, prezygotic mitosis, and the first and second postzygotic mitosis (189). There is also evidence of DNA synthesis occurring during meiotic prophase, and this could be interpreted as DNA repair synthesis ((141, 187). Thus, micronuclear localization suggests that these proteins may be playing roles in DNA mismatch repair rather than in meiosis.

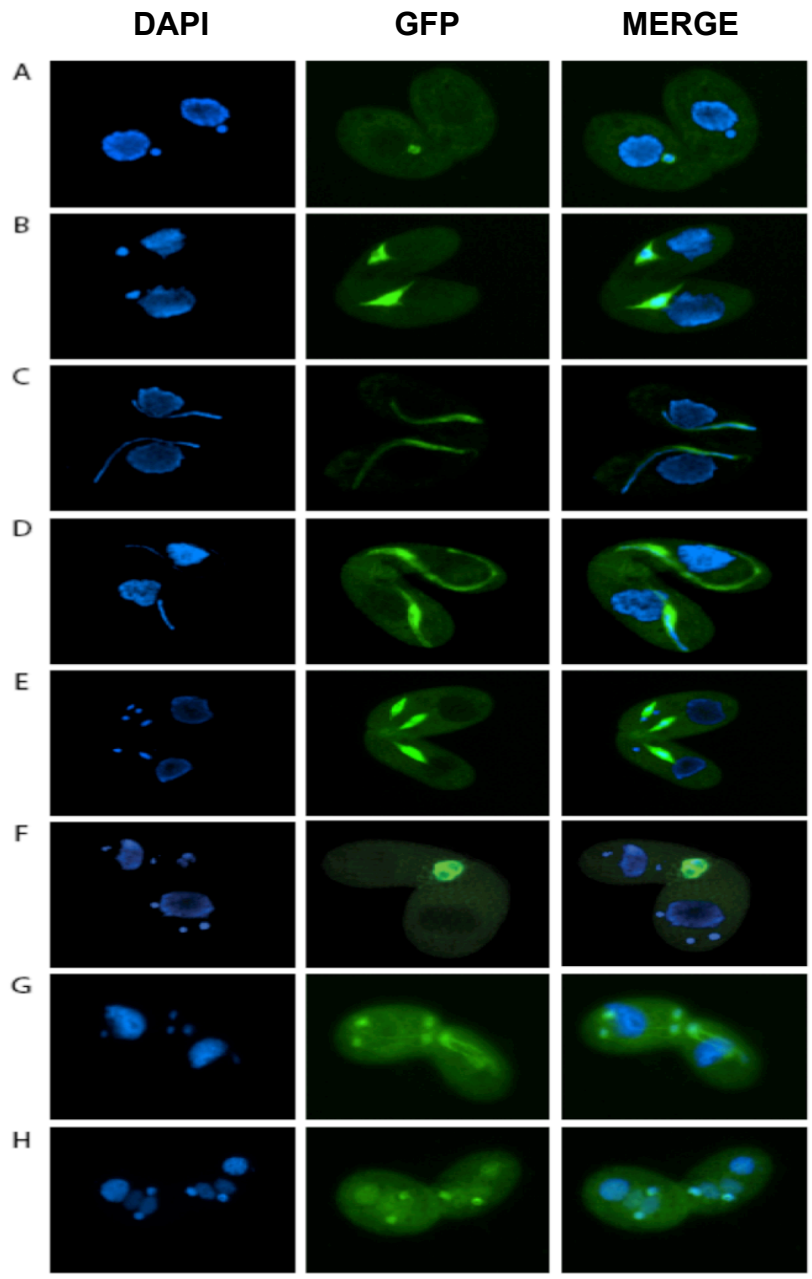


Figure 22. Localization of GFP-PMS2 protein in conjugation cells.

Wild-type cells were mated with a strain over-expressing GFP-tagged PMS2p. Localization is exclusively to the micronucleus during pairing (A), micronuclear elongation (B), micronuclear "crescent" phase (C and D), Meiosis I (bottom partner) and Meiosis II (top partner) (E) and pronuclear exchange (F). Localization remains in each mitotic product after the postzygotic mitotic divisions of the synkaryon (G), but eventually appears only in the new micronuclei during nuclear differentiation (H) . Left panels: DAPI; Center panels: GFP; Right panels: DAPI and GFP merge.

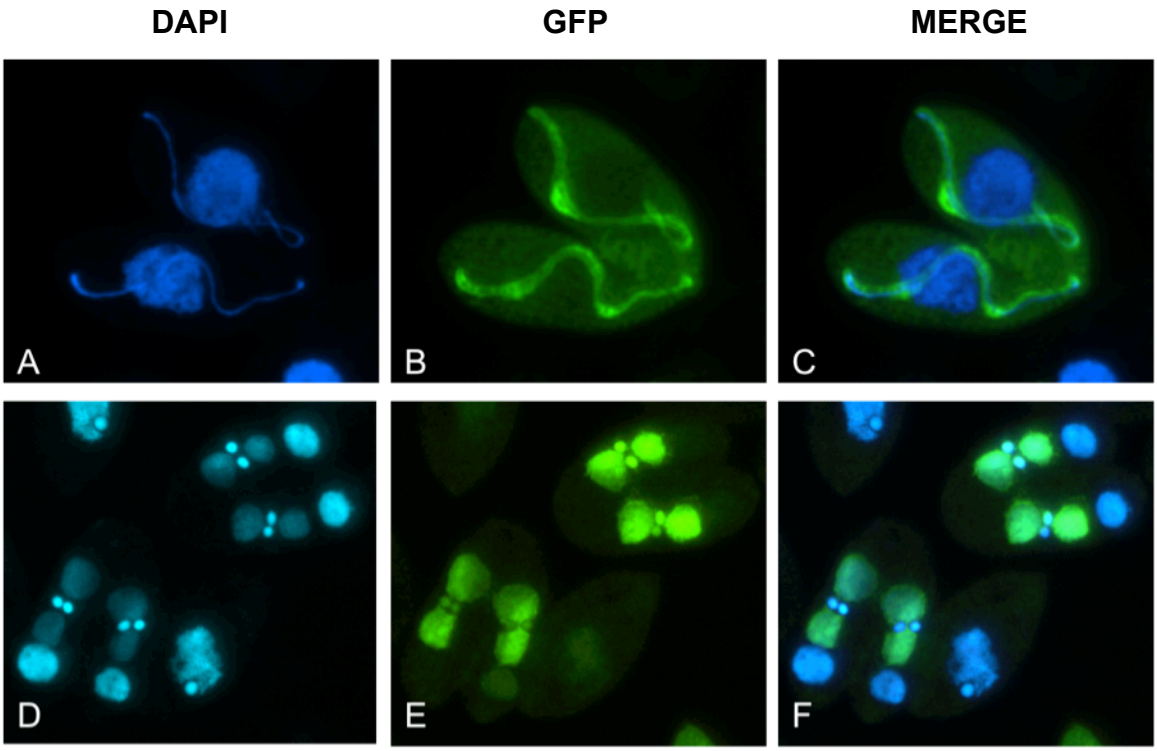


Figure 23. Localization of GFP-MSH6_1 proteins during "crescent" phase and anlagen development.

Wild-type cells were mated with a strain over-expressing GFP-tagged MSH6_1p. Localization during meiosis, and specifically "crescent" phase (A-C), is solely to the micronucleus. During the subsequent anlagen development (D-F), localization is both to the new micronuclei and new macronuclei. Left panels (A,D): DAPI; Center panels (B,E): GFP; Right panels (C,F): DAPI and GFP merge.

5.3.3 MMR and macronuclear development

Quantitative PCR studies of the MMR genes during conjugation revealed a bimodal pattern of expression. There was a considerable increase in transcript abundance during meiotic prophase and for some genes this was followed by a slight increase occurring around 8 hours, which corresponds to the period when postzygotic nuclei are completing

their final mitosis and beginning nuclear differentiation. This is also when degradation of the old macronucleus begins. Since duplicate, rather than triplicate, measurements were obtained for this data, it was important to establish the merit of these findings using available microarray data. Also, since our gene expression data ended at 9 hours into conjugation, we were interested in assessing the microarray data to determine whether levels of any of the MMR gene transcripts are altered during the later stages of conjugation (181). It indicated that all the MMR genes exhibited a bimodal pattern of expression with a second increase in transcript abundance during postzygotic development. This data suggests a role of the MMR proteins in late nuclear events during conjugation, including nuclear differentiation and macronuclear development.

As previously mentioned, during conjugation, gene expression from the new macronuclei begins near the end of postzygotic mitosis and the beginning of nuclear differentiation. Thus it is generally not possible to study phenotypes of macronuclear knockouts past this point. Interestingly, the Δ*MSH6_2* strain had an increased variability in pairing efficiency starting at around 8 hours post mixing and continuing until cell unpairing. The premature unpairing of cells at this point is indicative of a potential developmental block and suggests that the MSH6_2 proteins may be needed in some capacity for the transition from postzygotic mitosis to nuclear differentiation. Indeed, when the Δ*MSH6_2c1* strain was mated to wild-type B2086.2, there was a restoration of pairing efficiency during the later time points at which the KO cross showed a decrease. Preliminary analysis of Δ*MSH6_3* and Δ*MSH6_4* strains indicated that they may have a similar functional role during conjugation as that of the Δ*MSH6_2*,. Micronuclear

knockouts of these strains would be valuable for studies of later nuclear events and to discern a potential role for the proteins during conjugation. Localization data using GFP-tagged MMR proteins, indicates that these proteins have no distinct localization to nuclei, thus they may not be required in a repair capacity, assuming this data is not artifactual.

Concluding Remarks and Future Objectives

Examination of the macronuclear genome of *Tetrahymena* has revealed the presence of two MutL homologues, PMS2 and TML1, and five MutS homologues, MSH2 and MSH6_1 to 4. The unusual complement of MSH6s appears to be a ciliate specific occurrence and is reflective of an expansion of functional roles. Indeed, knockout data and GFP-localization data suggest that MSH6_2, MSH6_3 and MSH6_4 proteins are mainly functional during the intermediate stages of conjugation, during the postzygotic development of synkaryons, but they most likely aren't functional in mismatch repair, as evidenced by the lack of similar localization to that of MSH2. The MSH2 and MSH6_1 proteins are the canonical MMR proteins and are most similar to other eukaryotic MSH2 and MSH6 proteins. Coincidental localization of these two proteins and their similar phenotypes suggests that they most likely function together as a MutSα heterodimer in micronuclear mismatch repair with additional roles in macronuclear development not excluded, specifically during genome rearrangement and endoreplication. Similarly, the PMS2 and TML1 homologues most likely interact to form the MutLα heterodimer. The opposing localizations of the TML1 and PMS2 aside, these two homologues share similar phenotypes in knockouts, based on pairing efficiencies, and also lack micronuclear target as determined by the RMC assay. Thus it seems unlikely that both proteins would function independently yet still have similar phenotypes. Overall, these initial studies have suggested the MMR proteins of *Tetrahymena* may have limited roles in the regulation of meiotic recombination but may have accessory roles in the events leading up to and accompanying anlagen formation and macronuclear maturation during

conjugation. The principal functionality of these proteins appears to be during in the maintenance of micronuclear integrity during vegetative growth and in the transition from starvation to conjugation.

Macronuclear knockouts of the MMR genes have provided preliminary insight into the roles of these proteins during vegetative and conjugative development. Quantitative PCR studies and localization data have indicated that the MMR homologues may have important roles during the macronuclear developmental stage of conjugation. To better understand these roles it will be essential to create micronuclear knockout heterokaryons of the MMR genes (190). In these cells, the micronucleus contains a homozygous knockout of the gene of interest, while the macronucleus continues to express the wild-type alleles. When these heterokaryons conjugate, the old macronuclei degrade and the postzygotic nuclei differentiate into two micronuclei and two macronuclei in each partner. Because the MAC genome is derived from that of the MIC, it will also become a homozygous knockout for that gene. Thus the phenotype attributed to these knockouts can only be observed during the later stages of conjugation and in exconjugants. Additionally, since no phenotype is imposed upon the cells during vegetative growth, the accelerated ageing we observe in knockouts can be prevented.

The importance of MMR in vegetative growth and in the transition of cells into conjugation also needs further investigation. In my opinion, these periods are of upmost importance in terms of the faithful reproduction and transference of the micronuclear genome to progeny. It serves no benefit to *Tetrahymena* to allow micronuclear mutations

to accumulate and at the same time allow the macronucleus to maintain it's plasticity. It would be useful to extend the studies described here and follow MMR gene expression throughout vegetative growth. Unlike conjugation, vegetative growth is not synchronised among cells in culture, thus it is difficult to observe nuclear stages and perform and stage related analysis. This can be accomplished though through centrifugal elutriation (191, 192), allowing cells at different stages of the cell-cycle to be separated based on size. Additionally, a more thorough analysis of DNA content and chromosome integrity could be determined using FACS analysis of nuclei (90) and pulse-field gel electrophoresis, respectively. These experiments would allow us to determine whether MMR influences chromosome segregation during mitosis or if it perhaps even plays a role in preventing recombination between chromosomes. In yeast, the MMR proteins play an anti-recombination role in the prevention of recombination between divergent sequences and thus preventing deleterious genome rearrangements (193).

One of the key questions that requires investigation is: Are the putative *Tetrahymena* MMR homologues functional in DNA mismatch repair? The studies here provide some evidence of a role of MMR in the development of *Tetrahymena*, but a demonstration of repair function will only clarify these roles further. This can be accomplished by measuring the repair of DNA heteroduplexes with purified proteins in a reconstitution assays (194). Unfortunately, the genetic code of *Tetrahymena* is such that TAA and TAG encode for glutamine, but in most other organisms, such as *E. coli*, these are both stop codons. Thus in order to clone and purify proteins from *Tetrahymena* into *E. coli*, codons must changed. Alternatively, proteins can be purified directly from *Tetrahymena*.

The unique genetics and life cycle of *Tetrahymena* make it a novel organism to study the DNA mismatch repair system. These preliminary studies have provided useful and interesting information regarding the potential roles of the MMR proteins during both vegetative and conjugal development, There is still much to be learned about these putative proteins and their involvement in a wide variety of repair and mutation avoidance pathways. *Tetrahymena* has proven itself a worthy model for the study of multiple biological phenomena. I believe that with continued research, this will continue to remain the case for study of DNA mismatch repair and mutation avoidance in *Tetrahymena.*

Literature Cited

1. Crick,F. (1974) The double helix: a personal view. *Nature*, **248**, 766–769.

2. Watson,J.D. and Crick,F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737–738.

3. Wilkins,M.H.F., Stokes,A.R. and Wilson,H.R. (1953) Molecular structure of deoxypentose nucleic acids. *Nature*, **171**, 738–740.

4. Franklin,R.E. and Gosling,R.G. (1953) Molecular configuration in sodium thymonucleate. *Nature*, **171**, 740–741.

5. Friedberg,E.C. (2003) DNA damage and repair. *Nature*, **421**, 436–440.

6. Polosina,Y.Y. and Cupples,C.G. (2010) Wot the 'L-Does MutL do? *Mutat Res*, **705**, 228–238.

7. Polosina,Y.Y. and Cupples,C.G. (2010) MutL: conducting the cell's response to mismatched and misaligned DNA. *Bioessays*, **32**, 51–59.

8. Lynch,H.T., Lynch,P.M., Lanspa,S.J., Snyder,C.L., Lynch,J.F. and Boland,C.R. (2009) Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clin. Genet.*, **76**, 1–18.

9. Li,G.-M. (2008) Mechanisms and functions of DNA mismatch repair. *Cell Res*, **18**, 85–98.

10. Phillips,D.H. and Venitt,S. (2012) DNA and protein adducts in human tissues resulting from exposure to tobacco smoke. *Int. J. Cancer*, **131**, 2733–2753.

11. Rydberg,B. and Lindahl,T. (1982) Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction. *EMBO J*, **1**, 211–216.

12. Dizdaroglu,M. (2012) Oxidatively induced DNA damage: Mechanisms, repair and disease. *Cancer letters*, **327**, 26–47.

13. Kow,Y.W. (2002) Repair of deaminated bases in DNA. *Free Radic. Biol. Med.*, **33**, 886–893.

14. Lindahl,T. (1993) Instability and decay of the primary structure of DNA. *Nature*, **362**, 709–715.

15. Carell,T., Burgdorf,L.T., Kundu,L.M. and Cichon,M. (2001) The mechanism of action of DNA photolyases. *Curr Opin Chem Biol*, **5**, 491–498.

16. Benjdia,A. (2012) DNA photolyases and SP lyase: structure and mechanism of light-dependent and independent DNA lyases. *Curr. Opin. Struct. Biol.*, **22**, 711–720.

17. Parsons,J.L. and Dianov,G.L. (2013) Co-ordination of base excision repair and genome stability. *DNA Repair*, **12**, 326–333.

18. Schärer,O.D. (2003) Chemistry and biology of DNA repair. *Angew Chem Int Ed Engl*, **42**, 2946–2974.

19. Kamileri,I., Karakasilioti,I. and Garinis,G.A. (2012) Nucleotide excision repair: new tricks with old bricks. *Trends in Genetics*, **28**, 566–573.

20. Sonoda,E., Sasaki,M.S., Buerstedde,J.M., Bezzubova,O., Shinohara,A., Ogawa,H., Takata,M., Yamaguchi-Iwai,Y. and Takeda,S. (1998) Rad51-deficient vertebrate cells accumulate chromosomal breaks prior to cell death. *EMBO J*, **17**, 598–608.

21. van Gent,D.C.D., Hoeijmakers,J.H.J. and Kanaar,R.R. (2001) Chromosomal stability and the DNA double-stranded break connection. *Nat Rev Genet*, **2**, 196–206.

22. Chapman,J.R., Taylor,M.R.G. and Boulton,S.J. (2012) Playing the end game: DNA double-strand break repair pathway choice. *Mol Cell*, **47**, 497–510.

23. Lin,W.Y., Wilson,J.H. and Lin,Y. (2013) Repair of chromosomal double-strand breaks by precise ligation in human cells. *DNA Repair*, 10.1016/j.dnarep.2013.04.024.

24. Pitsikas,P., Polosina,Y.Y. and Cupples,C.G. (2009) Interaction between the mismatch repair and nucleotide excision repair pathways in the prevention of 5-azacytidine-induced CG-to-GC mutations in Escherichia coli. *DNA Repair*, **8**, 354–359.

25. Bertrand,P., Tishkoff,D.X., Filosi,N., Dasgupta,R. and Kolodner,R.D. (1998) Physical interaction between components of DNA mismatch repair and nucleotide excision repair. *Proc Natl Acad Sci USA*, **95**, 14278–14283.

26. Zhao,J., Jain,A., Iyer,R.R., Modrich,P.L. and Vasquez,K.M. (2009) Mismatch repair and nucleotide excision repair proteins cooperate in the recognition of DNA interstrand crosslinks. *Nucleic Acids Res*, **37**, 4420–4429.

27. Chu,Y.-L., Wu,X., Xu,Y. and Her,C. (2013) MutS homologue hMSH4: interaction with eIF3f and a role in NHEJ-mediated DSB repair. *Mol. Cancer*, **12**, 51.

28. Smith,J.A., Bannister,L.A., Bhattacharjee,V., Wang,Y., Waldman,B.C. and Waldman,A.S. (2007) Accurate homologous recombination is a prominent double-strand break repair pathway in mammalian chromosomes and is modulated by mismatch repair protein Msh2. *Mol Cell Biol*, **27**, 7816–7827.

29. Evans,E., Sugawara,N., Haber,J.E. and Alani,E. (2000) The Saccharomyces cerevisiae Msh2 mismatch repair protein localizes to recombination intermediates in

vivo. *Mol Cell*, **5**, 789–799.

30. Shahi,A., Lee,J.H., Kang,Y., Lee,S.H., Hyun,J.W., Chang,I.Y., Jun,J.Y. and You,H.J. (2011) Mismatch-repair protein MSH6 is associated with Ku70 and regulates DNA double-strand break repair. *Nucleic Acids Res*, **39**, 2130–2143.

31. Lin,Z., Nei,M. and Ma,H. (2007) The origins and early evolution of DNA mismatch repair genes--multiple horizontal gene transfers and co-evolution. *Nucleic Acids Res*, **35**, 7591–7603.

32. Iyer,R.R., Pluciennik,A., Burdett,V. and Modrich,P.L. (2006) DNA mismatch repair: functions and mechanisms. *Chem Rev*, **106**, 302–323.

33. Fukui,K. (2010) DNA mismatch repair in eukaryotes and bacteria. *J Nucleic Acids*, **2010**.

34. Cox,E.C. (1976) Bacterial mutator genes and the control of spontaneous mutation. *Annu Rev Genet*, **10**, 135–156.

35. Claverys,J.P. and Lacks,S.A. (1986) Heteroduplex deoxyribonucleic acid base mismatch repair in bacteria. *Microbiol Rev*, **50**, 133–165.

36. Fijalkowska,I.J.I., Schaaper,R.M.R. and Jonczyk,P.P. (2012) DNA replication fidelity in Escherichia coli: a multi-DNA polymerase affair. *FEMS Microbiol Rev*, **36**, 1105–1121.

37. Schaaper,R.M. and Dunn,R.L. (1987) Spectra of spontaneous mutations in Escherichia coli strains defective in mismatch correction: the nature of in vivo DNA replication errors. *Proc Natl Acad Sci USA*, **84**, 6220–6224.

38. Wu,J., Gu,L., Wang,H., Geacintov,N.E. and Li,G.M. (1999) Mismatch repair processing of carcinogen-DNA adducts triggers apoptosis. *Mol Cell Biol*, **19**, 8292–8301.

39. Bielas,J.H., Loeb,K.R., Rubin,B.P., True,L.D. and Loeb,L.A. (2006) Human cancers express a mutator phenotype. *Proc Natl Acad Sci USA*, **103**, 18238–18242.

40. Martín-López,J.V. and Fishel,R. (2013) The mechanism of mismatch repair and the functional analysis of mismatch repair defects in Lynch syndrome. *Fam Cancer*, 10.1007/s10689-013-9635-x.

41. Pukkila,P.J., Peterson,J., Herman,G., Modrich,P. and Meselson,M. (1983) Effects of high levels of DNA adenine methylation on methyl-directed mismatch repair in Escherichia coli. *Genetics*, **104**, 571–582.

42. Lu,A.L., Clark,S. and Modrich,P. (1983) Methyl-directed repair of DNA base-pair mismatches in vitro. *Proc Natl Acad Sci USA*, **80**, 4639–4643.

43. Jacobs-Palmer,E. and Hingorani,M.M. (2007) The effects of nucleotides on MutS-DNA binding kinetics clarify the role of MutS ATPase activity in mismatch repair. *J Mol Biol*, **366**, 1087–1098.

44. Acharya,S. (2008) Mutations in the signature motif in MutS affect ATP-induced clamp formation and mismatch repair. *Mol Microbiol*, **69**, 1544–1559.

45. Sancar,A. and Hearst,J.E. (1993) Molecular matchmakers. *Science*, **259**, 1415–1420.

46. Hall,M.C. and Matson,S.W. (1999) The Escherichia coli MutL protein physically interacts with MutH and stimulates the MutH-associated endonuclease activity. *J Biol Chem*, **274**, 1306–1312.

47. Robertson,A.B., Pattishall,S.R., Gibbons,E.A. and Matson,S.W. (2006) MutL-catalyzed ATP hydrolysis is required at a post-UvrD loading step in methyl-directed mismatch repair. *J Biol Chem*, **281**, 19949–19959.

48. Burdett,V., Baitinger,C., Viswanathan,M., Lovett,S.T. and Modrich,P. (2001) In vivo requirement for RecJ, ExoVII, ExoI, and ExoX in methyl-directed mismatch repair. *Proc Natl Acad Sci USA*, **98**, 6765–6770.

49. Koonin,E.V. (2005) Orthologs, paralogs, and evolutionary genomics. *Annu Rev Genet*, **39**, 309–338.

50. Marsischky,G.T., Filosi,N., Kane,M.F. and Kolodner,R. (1996) Redundancy of Saccharomyces cerevisiae MSH3 and MSH6 in MSH2-dependent mismatch repair. *Genes Dev*, **10**, 407–420.

51. Modrich,P. (2006) Mechanisms in eukaryotic mismatch repair. *J Biol Chem*, **281**, 30305–30309.

52. Ross-Macdonald,P. and Roeder,G.S. (1994) Mutation of a meiosis-specific MutS homolog decreases crossing over but not mismatch correction. *Cell*, **79**, 1069–1080.

53. 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.

54. Chi,N.W. and Kolodner,R.D. (1994) Purification and characterization of MSH1, a yeast mitochondrial protein that binds to DNA mismatches. *J Biol Chem*, **269**, 29984–29992.

55. Williamson,M.S., Game,J.C. and Fogel,S. (1985) Meiotic gene conversion mutants in Saccharomyces cerevisiae. I. Isolation and characterization of pms1-1 and pms1-2. *Genetics*, **110**, 609–646.

56. Baker,S.M., Bronner,C.E., Zhang,L., Plug,A.W., Robatzek,M., Warren,G., Elliott,E.A., Yu,J., Ashley,T., Arnheim,N., et al. (1995) Male mice defective in the

DNA mismatch repair gene PMS2 exhibit abnormal chromosome synapsis in meiosis. *Cell*, **82**, 309–319.

57. Prolla,T.A., Pang,Q., Alani,E., Kolodner,R.D. and Liskay,R.M. (1994) MLH1, PMS1, and MSH2 interactions during the initiation of DNA mismatch repair in yeast. *Science*, **265**, 1091–1093.

58. Li,G.M. and Modrich,P. (1995) Restoration of mismatch repair to nuclear extracts of H6 colorectal tumor cells by a heterodimer of human MutL homologs. *Proc Natl Acad Sci USA*, **92**, 1950–1954.

59. Cannavo,E., Marra,G., Sabates-Bellver,J., Menigatti,M., Lipkin,S.M., Fischer,F., Cejka,P. and Jiricny,J. (2005) Expression of the MutL homologue hMLH3 in human cells and its role in DNA mismatch repair. *Cancer Res*, **65**, 10759–10766.

60. Santucci-Darmanin,S., Neyton,S., Lespinasse,F., Saunières,A., Gaudray,P. and Paquis-Flucklinger,V. (2002) The DNA mismatch-repair MLH3 protein interacts with MSH4 in meiotic cells, supporting a role for this MutL homolog in mammalian meiotic recombination. *Hum Mol Genet*, **11**, 1697–1706.

61. Siehler,S.Y., Schrauder,M., Gerischer,U., Cantor,S., Marra,G. and Wiesmüller,L. (2009) Human MutL-complexes monitor homologous recombination independently of mismatch repair. *DNA Repair*, **8**, 242–252.

62. Yuan,F., Gu,L., Guo,S., Wang,C. and Li,G.-M. (2004) Evidence for involvement of HMGB1 protein in human DNA mismatch repair. *J Biol Chem*, **279**, 20935–20940.

63. Kadyrov,F.A., Dzantiev,L., Constantin,N. and Modrich,P. (2006) Endonucleolytic function of MutLalpha in human mismatch repair. *Cell*, **126**, 297–308.

64. Ghodgaonkar,M.M., Lazzaro,F., Olivera-Pimentel,M., Artola-Borán,M., Cejka,P., Reijns,M.A., Jackson,A.P., Plevani,P., Muzi-Falconi,M. and Jiricny,J. (2013) Ribonucleotides Misincorporated into DNA Act as Strand-Discrimination Signals in Eukaryotic Mismatch Repair. *Mol Cell*, **50**, 323–332.

65. Peled,J.U., Kuang,F.L., Iglesias-Ussel,M.D., Roa,S., Kalis,S.L., Goodman,M.F. and Scharff,M.D. (2008) The biochemistry of somatic hypermutation. *Annu. Rev. Immunol.*, **26**, 481–511.

66. Chahwan,R., Edelmann,W., Scharff,M.D. and Roa,S. (2012) AIDing antibody diversity by error-prone mismatch repair. *Semin. Immunol.*, **24**, 293–300.

67. Jun,S.-H., Kim,T.G. and Ban,C. (2006) DNA mismatch repair system. Classical and fresh roles. *FEBS Journal*, **273**, 1609–1619.

68. van Leewenhoeck,A. Observations, Communicated to the Publisher by Mr. Antony van Leewenhoeck, in a Dutch Letter of the 9th of Octob. 1676. Here English'd: concerning Little Animals by Him Observed in Rain-Well-Sea. and Snow Water; as

Also in Water Wherein Pepper Had Lain Infused. *Philosophical Transactions (1665-1678)*, **12**, 821–831.

69. Lwoff,A. (1923) Sur un Infusoire cilié homocaryote à vie libre. Son importance taxonomique. *CR Acad. Sci*, **177**, 910–912.

70. Soyer-Gobillard,M.-O. and Schrevel,J. (2003) André Lwoff (1902–1994), Nobel Prize of Medicine, as Protistologist. *Protist*, **154**, 455–468.

71. Blackburn,E.H. and Gall,J.G. (1978) A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. *J Mol Biol*, **120**, 33–53.

72. Greider,C.W. and Blackburn,E.H. (1985) Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell*, **43**, 405–413.

73. Kruger,K., Grabowski,P.J., Zaug,A.J., Sands,J., Gottschling,D.E. and Cech,T.R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, **31**, 147–157.

74. Brownell,J.E., Zhou,J., Ranalli,T., Kobayashi,R., Edmondson,D.G., Roth,S.Y. and Allis,C.D. (1996) Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell*, **84**, 843–851.

75. Gaertig,J., Gao,Y., Tishgarten,T., Clark,T.G. and Dickerson,H.W. (1999) Surface display of a parasite antigen in the ciliate Tetrahymena thermophila. *Nat Biotechnol*, **17**, 462–465.

76. Kiy,T. and Tiedtke,A. (1992) Mass cultivation of Tetrahymena thermophila yielding high cell densities and short generation times. *Applied microbiology and biotechnology*, **37**, 576–579.

77. Frankel,J. (2000) Cell biology of Tetrahymena thermophila. *Methods Cell Biol.*, **62**, 27–125.

78. Cassidy-Hanley,D., Bowen,J., Lee,J.H., Cole,E., VerPlank,L.A., Gaertig,J., Gorovsky,M.A. and Bruns,P.J. (1997) Germline and somatic transformation of mating Tetrahymena thermophila by particle bombardment. *Genetics*, **146**, 135–147.

79. Gaertig,J. and Gorovsky,M.A. (1992) Efficient mass transformation of Tetrahymena thermophila by electroporation of conjugants. *Proc Natl Acad Sci USA*, **89**, 9196–9200.

80. Eisen,J.A., Coyne,R.S., Wu,M., Wu,D., Thiagarajan,M., Wortman,J.R., Badger,J.H., Ren,Q., Amedeo,P., Jones,K.M., et al. (2006) Macronuclear Genome Sequence of the Ciliate Tetrahymena thermophila, a Model Eukaryote. *Plos Biol*, **4**, e286.

81. Orias,E. (2000) Toward Sequencing the Tetrahymena Genome: Exploiting the Gift of

Nuclear Dimorphism. *J Eukaryot Microbiol*, **47**, 328–333.

82. Allen,S.L. and Nanney,D.L. (1958) JSTOR: The American Naturalist, Vol. 92, No. 864 (May - Jun., 1958), pp. 139-160. *The American Naturalist*.

83. Orias,E. and Flacks,M. (1975) Macronuclear genetics of Tetrahymena. I. Random distribution of macronuclear genecopies in T. pyriformis, syngen 1. *Genetics*, **79**, 187–206.

84. Merriam,E.V. and Bruns,P.J. (1988) Phenotypic assortment in Tetrahymena thermophila: assortment kinetics of antibiotic-resistance markers, tsA, death, and the highly amplified rDNA locus. *Genetics*, **120**, 389–395.

85. Cervantes,M.D., Hamilton,E.P., Xiong,J., Lawson,M.J., Yuan,D., Hadjithomas,M., Miao,W. and Orias,E. (2013) Selecting One of Several Mating Types through Gene Segment Joining and Deletion in Tetrahymena thermophila. *Plos Biol*, **11**, e1001518.

86. Yao,M.C., Yao,C.H. and Monks,B. (1990) The controlling sequence for site-specific chromosome breakage in Tetrahymena. *Cell*, **63**, 763–772.

87. Yao,M.C. and Yao,C.H. (1994) Detection of circular excised DNA deletion elements in Tetrahymena thermophila during development. *Nucleic Acids Res*, **22**, 5702–5708.

88. Patil,N.S. and Karrer,K.M. (2000) A developmentally regulated deletion element with long terminal repeats has cis-acting sequences in the flanking DNA. *Nucleic Acids Res*, **28**, 1465–1472.

89. Fillingham,J.S. and Pearlman,R.E. (2004) Role of micronucleus-limited DNA in programmed deletion of mse2.9 during macronuclear development of Tetrahymena thermophila. *Eukaryotic Cell*, **3**, 288–301.

90. Davis,M.C., Ward,J.G., Herrick,G. and Allis,C.D. (1992) Programmed nuclear death: apoptotic-like degradation of specific nuclei in conjugating Tetrahymena. *Dev Biol*, **154**, 419–432.

91. Mpoke,S. and Wolfe,J. (1996) DNA digestion and chromatin condensation during nuclear death in Tetrahymena. *Exp Cell Res*, **225**, 357–365.

92. Cupples,C.G. and Miller,J.H. (1989) A set of lacZ mutations in Escherichia coli that allow rapid detection of each of the six base substitutions. *Proc Natl Acad Sci USA*, **86**, 5345–5349.

93. Cupples,C.G., Cabrera,M., Cruz,C. and Miller,J.H. (1990) A set of lacZ mutations in Escherichia coli that allow rapid detection of specific frameshift mutations. *Genetics*, **125**, 275–280.

94. Bielas,J.H. and Loeb,L.A. (2005) Quantification of random genomic mutations. *Nat Meth*, **2**, 285–290.

95. Wright,J.H., Modjeski,K.L., Bielas,J.H., Preston,B.D., Fausto,N., Loeb,L.A. and Campbell,J.S. (2011) A random mutation capture assay to detect genomic point mutations in mouse tissue. *Nucleic Acids Res*, **39**, e73.

96. Heinonen,T.Y. and Pearlman,R.E. (1994) A germ line-specific sequence element in an intron in Tetrahymena thermophila. *J Biol Chem*, **269**, 17428–17433.

97. Chilcoat,N.D., Melia,S.M., Haddad,A. and Turkewitz,A.P. (1996) Granule lattice protein 1 (Grl1p), an acidic, calcium-binding protein in Tetrahymena thermophila dense-core secretory granules, influences granule size, shape, content organization, and release but not protein sorting or condensation. *J Cell Biol*, **135**, 1775–1787.

98. Haddad,A. and Turkewitz,A.P. (1997) Analysis of exocytosis mutants indicates close coupling between regulated secretion and transcription activation in Tetrahymena. *Proc Natl Acad Sci USA*, **94**, 10675–10680.

99. Coyne,R.S., Thiagarajan,M., Jones,K.M., Wortman,J.R., Tallon,L.J., Haas,B.J., Cassidy-Hanley,D.M., Wiley,E.A., Smith,J.J., Collins,K., et al. (2008) Refined annotation and assembly of the Tetrahymena thermophila genome sequence through EST analysis, comparative genomic hybridization, and targeted gap closure. *BMC Genomics*, **9**, 562.

100. Corpet,F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res*, **16**, 10881–10890.

101. Hunter,S., Jones,P., Mitchell,A., Apweiler,R., Attwood,T.K., Bateman,A., Bernard,T., Binns,D., Bork,P., Burge,S., et al. (2012) InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res*, **40**, D306–12.

102. Punta,M., Coggill,P.C., Eberhardt,R.Y., Mistry,J., Tate,J., Boursnell,C., Pang,N., Forslund,K., Ceric,G., Clements,J., et al. (2012) The Pfam protein families database. *Nucleic Acids Res*, **40**, D290–301.

103. Letunic,I., Doerks,T. and Bork,P. (2012) SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res*, **40**, D302–5.

104. Pruitt,K.D., Tatusova,T., Brown,G.R. and Maglott,D.R. (2012) NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res*, **40**, D130–5.

105. Coyne,R.S., Hannick,L., Shanmugam,D., Hostetler,J.B., Brami,D., Joardar,V.S., Johnson,J., Radune,D., Singh,I., Badger,J.H., et al. (2011) Comparative genomics of the pathogenic ciliate Ichthyophthirius multifiliis, its free-living relatives and a host species provide insights into adoption of a parasitic lifestyle and prospects for disease control. *Genome Biol*, **12**, R100.

106. Magrane,M. and Consortium,U. (2011) UniProt Knowledgebase: a hub of integrated

protein data. *Database*, **2011**, bar009.

107. Swart,E.C., Bracht,J.R., Magrini,V., Minx,P., Chen,X., Zhou,Y., Khurana,J.S., Goldman,A.D., Nowacki,M., Schotanus,K., et al. (2013) The Oxytricha trifallax macronuclear genome: a complex eukaryotic genome with 16,000 tiny chromosomes. *Plos Biol*, **11**, e1001473.

108. Larkin,M.A., Blackshields,G., Brown,N.P., Chenna,R., McGettigan,P.A., McWilliam,H., Valentin,F., Wallace,I.M., Wilm,A., Lopez,R., et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.

109. Tamura,K., Peterson,D., Peterson,N., Stecher,G., Nei,M. and Kumar,S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*, **28**, 2731–2739.

110. Eckert,W.A., Kaffenberger,W., Krohne,G. and Franke,W.W. (1978) Introduction of hidden breaks during rRNA maturation and ageing in Tetrahymena pyriformis. *Eur J Biochem*, **87**, 607–616.

111. Chalker,D.L. and Yao,M.C. (2001) Nongenic, bidirectional transcription precedes and may promote developmental DNA deletion in Tetrahymena thermophila. *Genes Dev*, **15**, 1287–1298.

112. Mochizuki,K., Novatchkova,M. and Loidl,J. (2008) DNA double-strand breaks, but not crossovers, are required for the reorganization of meiotic nuclei in Tetrahymena. *Journal of Cell Science*, **121**, 2148–2158.

113. Gaertig,J., Gu,L., Hai,B. and Gorovsky,M.A. (1994) High frequency vector-mediated transformation and gene replacement in Tetrahymena. *Nucleic Acids Res*, **22**, 5391–5398.

114. Kahn,R.W., Andersen,B.H. and Brunk,C.F. (1993) Transformation of Tetrahymena thermophila by microinjection of a foreign gene. *Proceedings of the ….*

115. Bruns,P.J. and Cassidy-Hanley,D. (2000) Biolistic transformation of macro- and micronuclei. *Methods Cell Biol.*, **62**, 501–512.

116. Williams,N.E., Wolfe,J. and Bleyman,L.K. (2007) Long-Term Maintenance of Tetrahymena Spp. *Journal of Eukaryotic Microbiology*, **27**, 327–327.

117. Stuart,K.R. and Cole,E.S. (1999) Nuclear and cytoskeletal fluorescence microscopy techniques. *Methods Cell Biol.*, **62**, 291–311.

118. Stover,N.A., Krieger,C.J., Binkley,G., Dong,Q., Fisk,D.G., Nash,R., Sethuraman,A., Weng,S. and Cherry,J.M. (2006) Tetrahymena Genome Database (TGD): a new genomic resource for Tetrahymena thermophila research. *Nucleic Acids Res*, **34**, D500–3.

119. Stover,N.A., Punia,R.S., Bowen,M.S., Dolins,S.B. and Clark,T.G. (2012) Tetrahymena Genome Database Wiki: a community-maintained model organism database. *Database*, **2012**, bas007.

120. Csank,C., Taylor,F.M. and Martindale,D.W. (1990) Nuclear pre-mRNA introns: analysis and comparison of intron sequences from Tetrahymena thermophila and other eukaryotes. *Nucleic Acids Res*, **18**, 5133–5141.

121. Obmolova,G., Ban,C., Hsieh,P. and Yang,W. (2000) Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature*, **407**, 703–710.

122. Warren,J.J., Pohlhaus,T.J., Changela,A., Iyer,R.R., Modrich,P.L. and Beese,L.S. (2007) Structure of the human MutSalpha DNA lesion recognition complex. *Mol Cell*, **26**, 579–592.

123. Alani,E., Sokolsky,T., Studamire,B., Miret,J.J. and Lahue,R.S. (1997) Genetic and biochemical analysis of Msh2p-Msh6p: role of ATP hydrolysis and Msh2p-Msh6p subunit interactions in mismatch base pair recognition. *Mol Cell Biol*, **17**, 2436–2447.

124. Mendillo,M.L., Hargreaves,V.V., Jamison,J.W., Mo,A.O., Li,S., Putnam,C.D., Woods,V.L. and Kolodner,R.D. (2009) A conserved MutS homolog connector domain interface interacts with MutL homologs. *Proc Natl Acad Sci USA*, **106**, 22223–22228.

125. Lee,S.D., Surtees,J.A. and Alani,E. (2007) Saccharomyces cerevisiae MSH2-MSH3 and MSH2-MSH6 complexes display distinct requirements for DNA binding domain I in mismatch recognition. *J Mol Biol*, **366**, 53–66.

126. Junop,M.S., Obmolova,G., Rausch,K., Hsieh,P. and Yang,W. (2001) Composite active site of an ABC ATPase: MutS uses ATP to verify mismatch recognition and authorize DNA repair. *Mol Cell*, **7**, 1–12.

127. Clark,A.B., Valle,F., Drotschmann,K., Gary,R.K. and Kunkel,T.A. (2000) Functional interaction of proliferating cell nuclear antigen with MSH2-MSH6 and MSH2-MSH3 complexes. *J Biol Chem*, **275**, 36498–36501.

128. Kleczkowska,H.E., Marra,G., Lettieri,T. and Jiricny,J. (2001) hMSH3 and hMSH6 interact with PCNA and colocalize with it to replication foci. *Genes Dev*, **15**, 724–736.

129. Iyer,R.R., Pluciennik,A., Genschel,J., Tsai,M.-S., Beese,L.S. and Modrich,P. (2010) MutLalpha and proliferating cell nuclear antigen share binding sites on MutSbeta. *J Biol Chem*, **285**, 11730–11739.

130. Pluciennik,A., Dzantiev,L., Iyer,R.R., Constantin,N., Kadyrov,F.A. and Modrich,P. (2010) PCNA function in the activation and strand direction of MutLα endonuclease in mismatch repair. *Proc Natl Acad Sci USA*, **107**, 16066–16071.

131. Marass,F. and Upton,C. (2009) Sequence Searcher: A Java tool to perform regular expression and fuzzy searches of multiple DNA and protein sequences. *BMC Res Notes*, **2**, 14.

132. Cheng,J., Randall,A.Z., Sweredoski,M.J. and Baldi,P. (2005) SCRATCH: a protein structure and structural feature prediction server. *Nucleic Acids Res*, **33**, W72–6.

133. Culligan,K.M., Meyer-Gauen,G., Lyons-Weiler,J. and Hays,J.B. (2000) Evolutionary origin, diversification and specialization of eukaryotic MutS homolog mismatch repair proteins. *Nucleic Acids Res*, **28**, 463–471.

134. Garrison,E.M. and Arrizabalaga,G. (2009) Disruption of a mitochondrial MutS DNA repair enzyme homologue confers drug resistance in the parasite Toxoplasma gondii. *Mol Microbiol*, **72**, 425–441.

135. Wu,M., Allis,C.D., Sweet,M.T., Cook,R.G., Thatcher,T.H. and Gorovsky,M.A. (1994) Four distinct and unusual linker proteins in a mitotically dividing nucleus are derived from a 71-kilodalton polyprotein, lack p34cdc2 sites, and contain protein kinase A sites. *Mol Cell Biol*, **14**, 10–20.

136. Dutta,R. and Inouye,M. (2000) GHKL, an emergent ATPase/kinase superfamily. *Trends Biochem Sci*, **25**, 24–28.

137. Bergerat,A., de Massy,B., Gadelle,D., Varoutas,P.C., Nicolas,A. and Forterre,P. (1997) An atypical topoisomerase II from Archaea with implications for meiotic recombination. *Nature*, **386**, 414–417.

138. 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.

139. Pang,Q., Prolla,T.A. and Liskay,R.M. (1997) Functional domains of the Saccharomyces cerevisiae Mlh1p and Pms1p DNA mismatch repair proteins and their relevance to human hereditary nonpolyposis colorectal cancer-associated mutations. *Mol Cell Biol*, **17**, 4465–4473.

140. Doerder,F.P. and Debault,L.E. (1975) Cytofluorimetric analysis of nuclear DNA during meiosis, fertilization and macronuclear development in the ciliate Tetrahymena pyriformis, syngen 1. *Journal of Cell Science*, **17**, 471–493.

141. Allis,C.D., Colavito-Shepanski,M. and Gorovsky,M.A. (1987) Scheduled and unscheduled DNA synthesis during development in conjugating Tetrahymena. *Dev Biol*, **124**, 469–480.

142. Martindale,D.W., Allis,C.D. and Bruns,P.J. (1985) RNA and protein synthesis during meiotic prophase in Tetrahymena thermophila. *J Protozool*, **32**, 644–649.

143. Pfaffl,M.W. (2001) A new mathematical model for relative quantification in real-

time RT-PCR. *Nucleic Acids Res*, **29**, e45.

144. Livak,K.J. and Schmittgen,T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, **25**, 402–408.

145. Wang,Z. and Spadoro,J. (1996) Determination of target copy number of quantitative standards used in PCR-based diagnostic assays.

146. Pryor,R.J. and Wittwer,C.T. (2006) Real-Time Polymerase Chain Reaction and Melting Curve Analysis - Springer. *Clinical Applications of PCR*.

147. Abtahi,H., Sadeghi,M.R., Shabani,M., Edalatkhah,H., Hadavi,R., Akhondi,M.M. and Talebi,S. (2011) Causes of bimodal melting curve: Asymmetric guanine-cytosine (GC) distribution causing two peaks in melting curve and affecting their shapes. *African Journal of Biotechnology*, **10**, 10196–10203.

148. Tostesen,E., Jerstad,G.I. and Hovig,E. (2005) Stitchprofiles.uio.no: analysis of partly melted DNA conformations using stitch profiles. *Nucleic Acids Res*, **33**, W573–W576.

149. 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.

150. Tam,S.M., Samipak,S., Britt,A. and Chetelat,R.T. (2009) Characterization and comparative sequence analysis of the DNA mismatch repair MSH2 and MSH7 genes from tomato. *Genetica*, **137**, 341–354.

151. Horwath,M., Kramer,W. and Kunze,R. (2002) Structure and expression of the Zea mays mutS-homologs Mus1 and Mus2. *Theor Appl Genet*, **105**, 423–430.

152. Orias,E., Cervantes,M.D. and Hamilton,E.P. (2011) Tetrahymena thermophila, a unicellular eukaryote with separate germline and somatic genomes. *Res Microbiol*, **162**, 578–586.

153. Xiong,J., Feng,L., Yuan,D., Fu,C. and Miao,W. (2010) Genome-wide identification and evolution of ATP-binding cassette transporters in the ciliate Tetrahymena thermophila: A case of functional divergence in a multigene family. *BMC Evol Biol*, **10**, 330.

154. Holland,I.B. and Blight,M.A. (1999) ABC-ATPases, adaptable energy generators fuelling transmembrane movement of a variety of molecules in organisms from bacteria to humans. *J Mol Biol*, **293**, 381–399.

155. Zufall,R.A., McGrath,C.L., Muse,S.V. and Katz,L.A. (2006) Genome architecture drives protein evolution in ciliates. *Mol Biol Evol*, **23**, 1681–1687.

156. Foster,P.L. (2006) Methods for determining spontaneous mutation rates. *Meth Enzymol*, **409**, 195–213.

157. Kohler,S.W., Provost,G.S., Fieck,A., Kretz,P.L., Bullock,W.O., Sorge,J.A., Putman,D.L. and Short,J.M. (1991) Spectra of spontaneous and mutagen-induced mutations in the lacI gene in transgenic mice. *Proc Natl Acad Sci USA*, **88**, 7958–7962.

158. Brito,P.H., Guilherme,E., Soares,H. and Gordo,I. (2010) Mutation accumulation in Tetrahymena. *BMC Evol Biol*, **10**, 354.

159. Jin,Y.H., Clark,A.B., Slebos,R.J.C., Al-Refai,H., Taylor,J.A., Kunkel,T.A., Resnick,M.A. and Gordenin,D.A. (2003) Cadmium is a mutagen that acts by inhibiting mismatch repair. *Nat Genet*, **34**, 326–329.

160. Banerjee,S. and Flores-Rozas,H. (2005) Cadmium inhibits mismatch repair by blocking the ATPase activity of the MSH2-MSH6 complex. *Nucleic Acids Res*, **33**, 1410–1419.

161. Poovathingal,S.K., Gruber,J., Ng,L.F., Halliwell,B. and Gunawan,R. (2012) Maximizing signal-to-noise ratio in the random mutation capture assay. *Nucleic Acids Res*, **40**, e35.

162. Yao,M.-C. and Chao,J.-L. (2005) RNA-guided DNA deletion in Tetrahymena: an RNAi-based mechanism for programmed genome rearrangements. *Annu Rev Genet*, **39**, 537–559.

163. Li,J. and Pearlman,R.E. (1996) Programmed DNA rearrangement from an intron during nuclear development in Tetrahymena thermophila: molecular analysis and identification of potential cis-acting sequences. *Nucleic Acids Res*, **24**, 1943–1949.

164. Fillingham,J.S., Bruno,D. and Pearlman,R.E. (2001) Cis-acting requirements in flanking DNA for the programmed elimination of mse2.9: a common mechanism for deletion of internal eliminated sequences from the developing macronucleus of Tetrahymena thermophila. *Nucleic Acids Res*, **29**, 488–498.

165. Gorovsky,M.A., Yao,M.C., Keevert,J.B. and Pleger,G.L. (1975) Isolation of micro- and macronuclei of Tetrahymena pyriformis. *Methods Cell Biol.*, **9**, 311–327.

166. Carmody,M.W. and Vary,C.P. (1993) Inhibition of DNA hybridization following partial dUTP substitution. *BioTechniques*, **15**, 692–699.

167. Mühl,H., Kochem,A.-J., Disqué,C. and Sakka,S.G. (2010) Activity and DNA contamination of commercial polymerase chain reaction reagents for the universal 16S rDNA real-time polymerase chain reaction detection of bacterial pathogens in blood. *Diagn. Microbiol. Infect. Dis.*, **66**, 41–49.

168. Ehricht,R., Hotzel,H., Sachse,K. and Slickers,P. (2007) Residual DNA in

thermostable DNA polymerases – a cause of irritation in diagnostic PCR and microarray assays. *Biologicals*, **35**, 145–147.

169. Glushkov,S.A., Bragin,A.G. and Dymshits,G.M. (2009) Decontamination of polymerase chain reaction reagents using DEAE-cellulose. *Anal Biochem*, **393**, 135–137.

170. Sanchez,J.A., Pierce,K.E., Rice,J.E. and Wangh,L.J. (2004) Linear-after-the-exponential (LATE)-PCR: an advanced method of asymmetric PCR and its uses in quantitative real-time analysis. *Proc Natl Acad Sci USA*, **101**, 1933–1938.

171. Greaves,L.C., Beadle,N.E., Taylor,G.A., Commane,D., Mathers,J.C., Khrapko,K. and Turnbull,D.M. (2009) Quantification of mitochondrial DNA mutation load. *Aging Cell*, **8**, 566–572.

172. Rotskaya,U.N., Rogozin,I.B., Vasyunina,E.A., Malyarchuk,B.A., Nevinsky,G.A. and Sinitsyna,O.I. (2010) High frequency of somatic mutations in rat liver mitochondrial DNA. *Mutat Res*, **685**, 97–102.

173. Vermulst,M., Bielas,J.H., Kujoth,G.C., Ladiges,W.C., Rabinovitch,P.S., Prolla,T.A. and Loeb,L.A. (2007) Mitochondrial point mutations do not limit the natural lifespan of mice. *Nat Genet*, **39**, 540–543.

174. Nanney,D.L. (1974) Aging and long-term temporal regulation in ciliated protozoa. A critical review. *Mech Ageing Dev*, **3**, 81–105.

175. Wenkert,D. and Allis,C.D. (1984) Timing of the appearance of macronuclear-specific histone variant hv1 and gene expression in developing new macronuclei of Tetrahymena thermophila. *J Cell Biol*, **98**, 2107–2117.

176. Brieger,A., Plotz,G., Hinrichsen,I., Passmann,S., Adam,R. and Zeuzem,S. (2012) C-terminal fluorescent labeling impairs functionality of DNA mismatch repair proteins. *PLoS ONE*, **7**, e31863.

177. Chen,W. and Jinks-Robertson,S. (1999) The role of the mismatch repair machinery in regulating mitotic and meiotic recombination between diverged sequences in yeast. *Genetics*, **151**, 1299–1313.

178. Kirk,K.E., Christ,C., McGuire,J.M., Paul,A.G., Vahedi,M., Stuart,K.R. and Cole,E.S. (2008) Abnormal Micronuclear Telomeres Lead to an Unusual Cell Cycle Checkpoint and Defects in Tetrahymena Oral Morphogenesis. *Eukaryotic Cell*, **7**, 1712–1723.

179. Hoffmann,E.R. and Borts,R.H. (2004) Meiotic recombination intermediates and mismatch repair proteins. *Cytogenet Genome Res*, **107**, 232–248.

180. Surtees,J.A., Argueso,J.L. and Alani,E. (2004) Mismatch repair proteins: key regulators of genetic recombination. *Cytogenet Genome Res*, **107**, 146–159.

181. Miao,W., Xiong,J., Bowen,J., Wang,W., Liu,Y., Braguinets,O., Grigull,J., Pearlman,R.E., Orias,E. and Gorovsky,M.A. (2009) Microarray analyses of gene expression during the Tetrahymena thermophila life cycle. *PLoS ONE*, **4**, e4429.

182. Snowden,T., Acharya,S., Butz,C., Berardini,M. and Fishel,R. (2004) hMSH4-hMSH5 recognizes Holliday Junctions and forms a meiosis-specific sliding clamp that embraces homologous chromosomes. *Mol Cell*, **15**, 437–451.

183. Whitby,M.C. (2005) Making crossovers during meiosis. *Biochem Soc Trans*, **33**, 1451–1455.

184. Sym,M. and Roeder,G.S. (1994) Crossover interference is abolished in the absence of a synaptonemal complex protein. *Cell*, **79**, 283–292.

185. Lynn,A., Soucek,R. and Börner,G.V. (2007) ZMM proteins during meiosis: crossover artists at work. *Chromosome Res*, **15**, 591–605.

186. Wolfe,J., Hunter,B. and Adair,W.S. (1976) A cytological study of micronuclear elongation during conjugation in Tetrahymena. *Chromosoma*, **55**, 289–308.

187. Loidl,J. and Scherthan,H. (2004) Organization and pairing of meiotic chromosomes in the ciliate Tetrahymena thermophila. *Journal of Cell Science*, **117**, 5791–5801.

188. Lukaszewicz,A., Howard-Till,R.A. and Loidl,J. (2013) Mus81 nuclease and Sgs1 helicase are essential for meiotic recombination in a protist lacking a synaptonemal complex. *Nucleic Acids Res*, 10.1093/nar/gkt703.

189. Cole,E.S., Cassidy-Hanley,D., Hemish,J., Tuan,J. and Bruns,P.J. (1997) A mutational analysis of conjugation in Tetrahymena thermophila. 1. Phenotypes affecting early development: meiosis to nuclear selection. *Dev Biol*, **189**, 215–232.

190. Hai,B. and Gorovsky,M.A. (1997) Germ-line knockout heterokaryons of an essential alpha-tubulin gene enable high-frequency gene replacement and a test of gene transfer from somatic to germ-line nuclei in Tetrahymena thermophila. *Proc Natl Acad Sci USA*, **94**, 1310–1315.

191. Tang,L., Pelech,S.L. and Berger,J.D. (1994) A cdc2-Like Kinase Associated with Commitment to Division in Paramecium tetraurelia. *J Eukaryot Microbiol*, **41**, 381–387.

192. Marsh,T.C., Cole,E.S., Stuart,K.R., Campbell,C. and Romero,D.P. (2000) RAD51 is required for propagation of the germinal nucleus in Tetrahymena thermophila. *Genetics*, **154**, 1587–1596.

193. Harfe,B.D. and Jinks-Robertson,S. (2000) DNA mismatch repair and genetic instability. *Annu Rev Genet*, **34**, 359–399.

194. Dzantiev,L., Constantin,N., Genschel,J., Iyer,R.R., Burgers,P.M. and Modrich,P.

(2004) A defined human system that supports bidirectional mismatch-provoked excision. *Mol Cell*, **15**, 31–41.

Appendix A
MutS homologue protein accession numbers

Supplementary Table 1. NCBI and UniProtKB protein accession numbers of MutS homologues from representative species

Organism	NCBI Accession Number	UniProtKB Accession Number	Protein Name
Arabidopsis thaliana	NP_189075.2	Q84LK0	MSH1
	NP_566804.3	O24617	MSH2
	NP_194284.2	O65607	MSH3
	NP_193469.2	F4JP48	MSH4
	NP_188683.3	F4JEP5	MSH5
	NP_192116.1	O04716	MSH6
	NP_850630.1	Q9SMV7	MSH7
Aspergillus fumigatus	XP_747281.1	Q4WCE1	MSH1
	XP_754634.1	Q4WXK7	MSH2
	XP_749062.1	Q4WGB7	MSH3
	XP_749950.2	Q4WKM0	MSH4
	XP_752482.2	Q4WSY8	MSH5
	XP_751995.1	Q4WP77	MSH6
Bacillus subtilis	NP_389586.2	P49849	MutS
Bos taurus	NP_001029756.1	Q3MHE4	MSH2
	NP_001193184.1	----	MSH4
	NP_001192428.1	----	MSH5
	NP_001179666.1	E1B9Q4	MSH6
Caenorhabditis elegans	NP_491202.1	Q9TXR4	MSH2
	NP_495451.1	Q23405	MSH4
	NP_502531.1	Q19272	MSH5
	NP_491163.1	Q9N3T8	MSH6
Candida albicans	XP_717359.1	Q5A6Q9	MSH1
	XP_715402.1	Q5A102	MSH2
	XP_714452.1	Q59Y41	MSH3
	XP_718245.1	Q5A989	MSH4
	XP_720464.1	Q5AFY0	MSH5
	XP_722156.1	Q5AL33	MSH6
Chlorocebus aethiops	AAU85549.1	Q5XXB5	MSH2
Danio rerio	NP_998689.1	Q803R6	MSH2
	NP_001103184.1	A2CEA8	MSH3
	NP_878280.1	Q803S7	MSH6
Dictyostelium discoideum	XP_643371.1	Q552L1	MSH1
	XP_643399.1	Q553L4	MSH2
	XP_001134558.1	Q1ZXH0	MSH3
	XP_638826.1	Q54QB8	MSH4
	XP_638383.1	Q54P75	MSH5
	XP_647085.1	Q55GU9	MSH6
Drosophila melanogaster	NP_523565.2	P43248	MSH2
	NP_648755.1	Q9VUM0	MSH6
Escherichia coli	NP_417213.1	P23909	MutS
Homo sapiens	NP_000242.1	P43246	MSH2
	NP_002430.3	P20585	MSH3
	NP_002431.2	O15457	MSH4
	NP_079535.4	O43196	MSH5
	NP_000170.1	P52701	MSH6
Ichthyophthirius multifiliis	EGR30181.1	G0QX73	MSH2
	EGR27216.1	G0R5P3	MSH6
	EGR29225.1	G0QZW3	MSH6
	EGR32403.1	G0QQV4	MSH6
Mus musculus	NP_032654.1	P43247	MSH2
	NP_034959.2	P13705	MSH3
	NP_114076.1	Q99MT2	MSH4
	NP_038628.2	Q9QUM7	MSH5
	NP_034960.1	P54276	MSH6
Paramecium tetraurelia	XP_001347169.1	Q6BFM8	MSH2
Plasmodium falciparum	XP_001348428.1	Q8ILI9	MSH2
Rattus norvegicus	NP_112320.1	P54275	MSH2
	NP_001178886.1	F1LQM8	MSH3
	NP_001099947.2	----	MSH4

Supplementary Table 1 cont. NCBI and UniProtKB protein accession numbers of MutS homologues from representative species

Organism	NCBI Accession Number	UniProtKB Accession Number	Protein Name
	NP_997701.2	Q6MG62	MSH5
Saccharomyces cerevisiae	NP_011988.1	P25846	MSH1
	NP_014551.1	P25847	MSH2
	NP_010016.2	P25336	MSH3
	NP_116652.1	P40965	MSH4
	NP_010127.1	Q12175	MSH5
	NP_010382.3	Q03834	MSH6
***Salmonella* Typhimurium**	NP_461830.1	P0A1Y0	MutS
Schizosaccharomyces pombe	NP_593649.2	O13921	MSH1
	XP_001713136.1	O74773	MSH2
	NP_593952.1	----	MSH3
	NP_588344.1	O74502	MSH6
Solanum lycopersicum	AAX53097.1	Q1XBQ6	MSH1
	NP_001234067.1	D6QY20	MSH2
	ADG85113.1	D6QY21	MSH7
Staphylococcus aureus	AAP87090.1	P0C1S1	MutS
Streptococcus pyogenes	YP_061143.1	Q5X9F3	MutS
Tetrahymena thermophila	ABK35677.1	A0MNQ6	MSH2
	ADQ26782.1	E5KKV6	MSH6_1
	ABK35676.1	A0MNQ5	MSH6_2
	ABK35678.1	A0MNQ7	MSH6_3
	ADQ26783.1	E5KKV7	MSH6_4
Thermus thermophilus	YP_004929.1	P61671	MutS
Toxoplasma gondii	ACI13852.1	B6DST1	MSH1
	XP_002365253.1	B6KBR9	MSH2
	XP_002368458.1	B6KKT5	MSH6
Trypanosoma brucei	AAK08648.1	Q9BLY3	MSH2
	AAN08917.1	Q8ISC8	MSH3
	AAK51796.1	Q967Z1	MSH8
Trypanosoma cruzi	XP_819877.1	Q4DZP7	MSH2
	XP_820278.1	Q4E0X8	MSH3
	XP_815283.1	Q4DLP2	MSH4
	XP_809831.1	Q4D633	MSH5
	XP_821503.1	Q4E4A8	MSH6
Xenopus laevis	S53609	----	MSH2
	NP_001089247.1	Q5FWN5	MSH6
Zea mays	NP_001105898.1	Q1XBQ7	MSH1
	NP_001146301.1	Q9XGC9	MSH2
	ADT92190.1	E7DDV5	MSH3
	AAF35250.1	Q9M5L9	MSH6
	CAB42555.1	Q9XGD0	MSH7

Appendix B
MutL homologue protein accession numbers

Supplementary Table 2. NCBI and UniProtKB protein accession numbers of MutL homologues from representative species.

Organism	NCBI Accession Number	UniProtKB Accession Number	Protein Name
Arabidopsis thaliana	NP_567345.2	Q9ZRV4	MLH1
	NP_195277.5	F4JN26	MLH3
	NP_567236.1	Q941I6	PMS1
Aspergillus fumigatus	XP_753480.1	Q4WVD4	MLH1
	XP_752177.1	Q4WNP5	MLH3
	XP_755674.1	Q4X0H9	PMS1
Bacillus subtilis	NP_389587.1	P49850	MutL
Bos taurus	NP_001069462.2	F1MPG0	MLH1
	XP_003586644.1	----	MLH3
	NP_001071449.1	A0JNP9	PMS1
	NP_001192867.1	----	PMS2
Caenorhabditis elegans	NP_499796.2	Q9XU10	MLH1
	NP_505933.1	G5EFG5	PMS2
Candida albicans[a]	XP_712766.1	Q59SL0	MLH1
	XP_715055.1	Q59ZT5	MLH3
	Orf19.1605p	----	PMS1
Chlorocebus aethiops	----	----	----
Danio rerio	NP_956953.1	Q6PFL1	MLH1
	XP_696739.3	----	MLH3
	NP_958476.2	Q8JFR9	PMS1
	XP_693648.4	----	PMS2
Dictyostelium discoideum	XP_637285.1	Q54KD8	MLH1
	XP_638891.4	Q54QI0	MLH3
	XP_638844.1	Q54QA0	PMS1
Drosophila melanogaster	NP_477022.1	A1Z7C1	MLH1
	NP_477023.1	A1ZA03	PMS2
Escherichia coli	NP_418591.1	P23367	MutL
Homo sapiens	NP_000240.1	P40692	MLH1
	NP_001035197.1	Q9UHC1	MLH3
	NP_000525.1	P54277	PMS1
	NP_000526.1	P54278	PMS2
Ichthyophthirius multifiliis	EGR27058.1	G0R635	PMS2
Mus musculus	NP_081086.2	Q9JK91	MLH1
	NP_780546.1	Q68FG1	MLH3
	NP_705784.1	Q8K119	PMS1
	NP_032912.2	B9EJ22	PMS2
Paramecium tetraurelia	XP_001441402.1	A0CT88	MLH1
	XP_001461612.1	A0EFZ8	PMS2
Plasmodium falciparum	XP_001347855.1	Q8IIJ0	MLH1
	XP_001349161.1	Q8IBJ3	PMS1
Rattus norvegicus	NP_112315.1	P97679	MLH1
	NP_001101513.1	D4ADG4	MLH3
	NP_001009535.1	Q6P7D0	PMS1
	AAI60858.1	B1H246	PMS2
Saccharomyces cerevisiae	NP_013890.1	P38920	MLH1
	NP_013135.1	Q07980	MLH2
	NP_015161.1	Q12083	MLH3
	NP_014317.4	P14242	PMS1
***Salmonella* Typhimurium**	NP_463220.1	P14161	MutL
Schizosaccharomyces pombe	NP_596199.1	Q9P7W6	MLH1
	NP_594417.1	P54280	PMS1
Solanum lycopersicum	NP_001234641.1	A5H619	MLH1
Staphylococcus aureus	YP_499806.1	Q93T05	MutL
Streptococcus pyogenes	YP_061121.1	Q5X9H5	MutL
Tetrahymena thermophila	ABK35675.1	A0MNQ4	TML1
	ABK35674.1	A0MNQ3	PMS2
Thermus thermophilus	YP_144589.1	Q9RA54	MutL
Toxoplasma gondii	XP_002368728.1	B6KLX8	MLH1
	XP_002369003.1	B6KMD0	PMS2
Trypanosoma brucei	AAK29067.1	Q9BIX4	MLH1
	AAK08649.1	Q9BLY2	PMS1
Trypanosoma cruzi	XP_814087.1	Q4DI77	MLH1
	XP_813339.1	Q4DG46	PMS1
Xenopus laevis	NP_001090545.1	A0AUU1	MLH1
	NP_001085796.2	Q6GNZ4	MLH3
	NP_001079545.1	Q7ZXV9	PMS1
Zea mays	----	----	----

[a] ***Candida albicans* PMS1** sequence was obtained from the *Candida* Genome Database (www.candidagenome.org)

Appendix C
Protozoan MSH6 protein accession numbers

Supplementary Table 3. NCBI and UniProtKB protein accession numbers of MSH6 homologues from representative protists.

Organism	Accession Number[a]	UniProtKB Accession Number	Protein Name[b]	Protein Size (aa)
Babesia bovis	XP_001609092.1	A7AWU6	MSH6	1313
Cryptosporidium parvum	XP_001388418.1	A3FQP9	MSH6	1242
Dictyostelium discoideum	XP_647085.1	Q55GU9	MSH6	1260
Entamoeba histolytica	XP_651951.1	C4M4T8	MSH6	934
Ichthyophthirius multifiliis	EGR29225.1	G0QZW3	MSH6_1	1077
	EGR27216.1	G0R5P3	MSH6_3	690
	EGR32403.1	G0QQV4	MSH6_4	1002
Leishmania major	XP_001686770.1	Q4Q1M8	MSH8	1014
Monosiga brevicollis	XP_001749687.1	A9VAJ6	MSH6	1131
Oxytricha trifallax	EJY65139.1	J9HMZ5	MSH6 (A)	1425
	EJY74495.1	J9IQT1	MSH6 (B)	1292
	EJY65086.1	J9EG13	MSH6 (C)	1185
	EJY82649.1	J9ISL4	MSH6 (D)	1348
Paramecium tetraurelia	XP_001455540.1	A0DYM6	MSH6 (A)	1106
	XP_001425815.1	A0BIQ0	MSH6 (B)	1108
	XP_001451767.1	A0DMV3	MSH6 (C)	1111
	XP_001459445.1	A0E9T1	MSH6 (D)	1084
	XP_001426119.1	A0BJK4	MSH6 (E)	1103
Plasmodium falciparum	XP_001351613.1	Q8I447	MSH6	1350
Tetrahymena borealis	EI9_03750	--	MSH6_1	1231
	EI9_00176	--	MSH6_2	2230
	EI9_14711	--	MSH6_3	1254
	EI9_05062	--	MSH6_4	1388
Tetrahymena elliotti	EI7_09598	--	MSH6_1	1235
	EI7_01503	--	MSH6_2	1137
	EI7_10743	--	MSH6_3	1251
	EI7_02303	--	MSH6_4	1389
Tetrahymena malaccensis	EIA_10158	--	MSH6_1	1232
	EIA_15108	--	MSH6_2	1136
	EIA_04644	--	MSH6_3	1254
	EIA_03957	--	MSH6_4	1392
Tetrahymena thermophila	ADQ26782.1	E5KKV6	MSH6_1	1232
	ABK35676.1	A0MNQ5	MSH6_2	1134
	ABK35678.1	A0MNQ7	MSH6_3	1257
	ADQ26783.1	E5KKV7	MSH6_4	1289
Thalassiosira pseudonana	XP_002288822.1	B8BXS4	MSH6	1099
Theileria parva	XP_765712.1	Q4N9C9	MSH6	1160
Toxoplasma gondii	XP_002368458.1	B6KKT5	MSH6	1607
Trichomonas vaginalis	XP_001322723.1	A2EA54	MSH6	1057
Trypanosoma brucei	AAK51796.1	Q967Z1	MSH8	997
Trypanosoma cruzi	XP_821503.1	Q4E4A8	MSH6	1002

[a] Accession numbers for *T. borealis, T. elliotti,* and *T. malaccensis* refer to Gene ID's belonging to the Broad Institute *Tetrahymena* Comparative Database (http://www.broadinstitute.org).

[b] Protein names for *O. trifallax* and *P. tetraurelia* contain the suffixes A, B, C, D and E only to distinguish between putative MSH6 paralogues.
Protein names given for *I. multifiliis, T. borealis, T. elliotti and T. malaccensis* are based on homology to *T. thermophila* MSH6's.

Appendix D
Protein sizes (aa) of MutS homologues from representative species

Organism	MutS	MSH1	MSH2	MSH3	MSH4	MSH5	MSH6	MSH7	MSH8
Arabidopsis thaliana	--	1118	937	1081	792	807	1324	1109	--
Aspergillus fumigatus	--	1018	940	1123	840	830	1213	--	--
Bacillus subtilis	858	--	--	--	--	--	--	--	--
Bos taurus	--	--	934	--	934	832	1360	--	--
Caenorhabditis elegans	--	--	849	--	842	1369	1186	--	--
Candida albicans	--	923	873	1037	803	623	1214	--	--
Chlorocebus aethiops	--	--	933	--	--	--	--	--	--
Danio rerio	--	--	936	1083	--	--	1369	--	--
Dictyostelium discoideum	--	898	937	1428	1041	880	1260	--	--
Drosophila melanogaster	--	--	917	--	--	--	1190	--	--
Escherichia coli	853	--	--	--	--	--	--	--	--
Homo sapiens	--	--	934	1137	936	822	1360	--	--
Ichthyophthirius multifiliis	--	--	705p	--	--	--	--	--	--
Mus musculus	--	--	935	1095	958	833	1358	--	--
Paramecium tetraurelia	--	--	794	--	--	--	--	--	--
Plasmodium falciparum	--	--	811	--	--	--	1350	--	--
Rattus norvegicus	--	--	933	1105	911	830	1361	--	--
Saccharomyces cerevisiae	--	959	964	1018	878	901	1242	--	--
Salmonella Typhimurium	855	--	--	--	--	--	--	--	--
Schizosaccharomyces pombe	--	941	982	1004	--	--	1254	--	--
Solanum lycopersicum	--	1124	943		--	--	--	782	--
Staphylococcus aureus	872	--	--	--	--	--	--	--	--
Streptococcus pyogenes	851	--	--	--	--	--	--	--	--
Tetrahymena thermophila	--	--	see below	--	--	--	see below	--	--
Thermus thermophilus	811	--	--	--	--	--	--	--	--
Toxoplasma gondii	--	2163	936	--	--	--	1607	--	--
Trypanosoma brucei	--	--	951	937	--	--	--	--	997
Trypanosoma cruzi	--	--	960	942	1264	784	1002	--	--
Xenopus laevis	--	--	933	--	--	--	1340	--	--
Zea mays	--	1131	942	981	--	--	629	1184	--

	MSH2	MSH6_1	MSH6_2	MSH6_3	MSH6_4
	813	**1232**	**1134**	**1257**	**1389**
Tetrahymena thermophila					

p = partial sequence

Appendix E
Protein sizes (aa) of MutL homologues from representative species

Organism	MutL	MLH1	MLH2	MLH3	PMS1	PMS2
Arabidopsis thaliana	--	737	--	1169	--	923
Aspergillus fumigatus	--	709	--	916	--	1044
Bacillus subtilis	627	--	--	--	--	--
Bos taurus	--	758	--	1460	932	864
Caenorhabditis elegans	--	758	--	--	--	805
Candida albicans	--	717	--	636	--	910
Chlorocebus aethiops	--	--	--	--	--	--
Danio rerio	--	724	--	1164	896	849
Dictyostelium discoideum	--	884	--	1658	--	1022
Drosophila melanogaster	--	664	--	--	--	899
Escherichia coli	615	--	--	--	--	--
Homo sapiens	--	756	--	1453	932	862
Ichthyophthirius multifiliis	--	--	--	--	--	549
Mus musculus	--	760	--	1411	917	859
Paramecium tetraurelia	--	623	--	--	--	685
Plasmodium falciparum	--	1016	--	--	--	1330
Rattus norvegicus	--	757	--	1442	919	853
Saccharomyces cerevisiae	--	769	695	715	--	873
Salmonella Typhimurium	618	--	--	--	--	--
Schizosaccharomyces pombe	--	684	--	--	--	794
Solanum lycopersicum	--	730	--	--	--	--
Staphylococcus aureus	669	--	--	--	--	--
Streptococcus pyogenes	673	--	--	--	--	--
Tetrahymena thermophila	--	see below	--	--	--	see below
Thermus thermophilus	545	--	--	--	--	--
Toxoplasma gondii	--	1219	--	--	--	1687
Trypanosoma brucei	--	887	--	--	--	788
Trypanosoma cruzi	--	864	--	--	--	774
Xenopus laevis	--	750	--	1305	925	--
Zea mays	--	--	--	--	--	--

Organism	TML1
Tetrahymena thermophila	**756**

Organism	PMS2
Tetrahymena thermophila	**946**

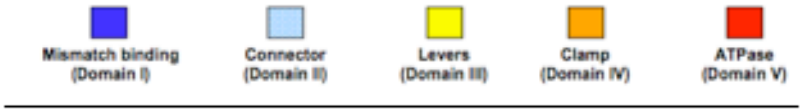
Appendix F
MSH4, MSH5 and MLH3 predicted protein sequence analysis

Predicted MSH4 (1537aa; TTHERM_00857890; XP_001021931.1)

MLNRKKNFRAHYSNKSSFSNTDDKGLEKSINQDFITPKNGFEYLKFNMKKHQNQIQSLDNSFQSIQQKNLIQELHQNFYSSNQGQFQYI
TPLLKQNGIDQHFNQTNNNYQTNSSNQMNSLQQQSNEQIHQNGFDRGTFPINKYFQFQNISFPFINGQSSSGLLNQINDSNYIIDKSFS
SLQQTNNQDVIKRKLSNFEEITPQMIENAYSNSNNFDSDLIKFQNEFPSLNQNSLYFDQNINTKCLNDIHNNQFQNKFQESNYQKDQKF
FFNFPSQLCDNQYEIDQKTLDFQQNQNNYIFFKQKQDQLNQALLDQSYNKNKGLYFSQNFNNNFIGEESLSQKSDQLDQFNKDQQNYFE
TDNIQQDQDISQVNFINDQIECKVNSQRDIFQINQDSQQQKNFNCFIEQVGYPNNQFNQNFNEIKQNFQNNQIYNLENDDQDQLDEFQN
QIQQQQNNEEVQNQNRIKHQCINQNEQFQINQISIDSVKPSQNNLLFNSTKSQKSSSKKQSPYQSLSENKLSFYIDSIKSQFKGKQFFN
SQKNNNEKQIQDSNLKDNQNNQFDYQKEEEQTEPKQNIFISKKEQFFSEQQNIIKHKNIEKNKILVFTDIAPKPVKDTSNQFQSHKDLK
KNHFSIAEEQEEYDDDDNDKNDDGEEEQEQEEEESKVQDILNQRDKQKMEEFDSIYNEFNNGLKQKLSDQIKINQLNQTNMFGQIDNHN
YKQQAELQESIFSSNKYQIDKRLNSNQMHNSKGMQKNDTKINSNKTQLQLFSKQVKKRKEINYSEDKLMQHIVAIIENHANQVGFAAFN
VQTGQISISQIVDRLTYLNTTSTIFAFYPLEIIFSNTQSSSYLVKILQQQLPLTKPTSVKRNYFDETKGESILKKKSSNLIQKLDNLYV
ALASLMALINHIEVTQELYLLKEQLQIGFIQMDSILVMDHKTALDLELLIQQLSQNEKGSLISLFKIQTVAGKRLLRSNLLQPLNDLKS
LQDRQSAVFEFTMNQDSILSQLQLQLQIIDMQNSYIKELKNLLCDQQLEQIKESIKQNLDLSYLEKIQNQQQHKSHDLIYFILKAEKDN
LLEIERTKMVSIQKYANDIFCEYQNKLIGYASLELKENENRGYYLIARKAKKQSQNINQKKIEELIEDNLIEVNDNKIGKSSQITFVTQ
IFAELSNNFYYSFSSEETPLCLPKLINEEINLLVLDEFYHPLLVYNNKDQWRRKSSNQNQQDESFGHSNVYIQKNSICISDLNNLQLFL
GQQNCGKSILIKSIAIIQIMAQVGCYVPAKNGVISLKNLILSKFYTTDSIEEKQSSFTSELQQLDFIIKSKASKDQQRGCLSVNQALSL
SKSKDKTIILIDEFCSSSQYEDNISISLAFLEEISICFPNSFLLCASKYYELVQMTQFYHNFQYFFINQDYQLKELNENDQENQMIDFS
EANKILYECLNPILLKYIITNQFENSQANDNQLLKTDQLHVQNKTLNKKIINDIVKQFVKIKDKEATFKANVNSKGFQIICFFFGVNEN
QKRSKCQTKMVFIITQRCLKERAV

Predicted MSH5 (1081aa; TTHERM_00763040; XP_001025362.1)

MIREGIVQELDDLLKIYDNLDEILSYYNSQEIIKLTNNNSNFNSQMMIQYFPQLGFFITIQNSNQFNNQVQVDDKVNFSQNKSYQDQDG
KQLDDDEVEGILKYNQILIYQSYYNQLKIIEKQNQSQQNLKRTRSSNSLSSKNQGIENSQNLQQQYCLTNDYSYCFTLQDALLFKNDLC
NQLDLKYGDIQSRIIDIQRDIILLYSEEINAAQEFISDLDLFFCLYQAAEVYNLRKPKLTNDSSLVFQEMRNILTEITIGEKQFISNSF
CVENFAQYMKNQKIQTNNFLDSQFSNRFTFITGPNYAGKSVFLKSIGMITYLAHIGSFIPCRFARIGMIDNIFVNTCQHDSVLKFNGES
AQQLQFINYILNQATNRTLILIDEFGKNFRLNDSLNLNYSLVKCLTEDTFQGLNEKITQKQHPSLQQEDNDQNIFYQYQTDNSQVENID
EIKQQNLRGLNEFQFIHQDIQNVNLMEQYNKIPITLMASHLSQLVNLCGLRENYIIRFAQMEVLIQQGKNVCKLNELKENINAMQKIVH
VYKLNKGLSLKSCSSLVAKNIFRDNEYLIHRINEINQFILEGKFQVSPQQNVEQIHLEKVQKVINILNSLKESLFSGIPTNFLAQHHQQ
LRARSVGKAVKSFTGKAKSTAQEEIAEVQDFKEELLEDQKKFIIQKEVEQERLKQQYEQQQFKTVEQQIENQNLDVTRHELIQKTKQRHD
QKAKQDASKEKQSNFDDLSNVRYYDEAQLTRRIRQSKNGESLFKLYNAEKFTYTTKNLVHTLTKYTQLVQQNKSNRLPKLDRRFMSLLM
SLQSNTNMLQNLEKSMILWSLCKLNMQEEIVEKLIKESVDNIKQLTFKQISILIWVMAKFSYKQTNQLNSVCLEVGLRLDDPSVKICTQ
NEIKKPTSQKKEEELDFEDLEQALELPEEKENIIEPQSLSMFLWSLGKLGVYDEDLMKSALNVFEGFCEMLQKKADQLKEENVLTLKTL
IVCLCRLKVGEQEFILDLAKMIINKTNFKKNPMSPRFMSEILYSLSKRYIYDEKIYNQMASNTIQVASSMNIKDCSESLYAFAKFKSLS
KLEVSRIKKQIKY



Predicted MLH3 (319aa; TTHERM_01044360, XP_001030620.1)

MISDYVFNQTYATIRSSKNSDNQKKRTSYIQTLSLKNDFKQKILENDYQDISLADRNQNFNQQIQIFSKQSNEEVQQISQVSKLDRTIF
EDIEIFGNCNNKVIICFNQQKGMLFGLDQHAIHERIRYEYFCNQFKASAFCLQYKQSRNEIDSKCINLSQRNKDKQFPSILWFDQTRTN
CLELDAYIFQRLQKNVDKLSQFKIQVVKIQNINQNKFEVYLWPQLYILNKPITYDLKFIDSILNSELGHIPNTIDEIIMSKACKGAIKF
NEELNQNQMDMLIKNIKLCEFFFYCVHGRSSIHPFFSLEIQDVCQIQKNYQI

Predicted MSH4, MSH5 and MLH3 homologues in *Tetrahymena thermophila*. Sequences were obtained from BLAST analysis of the *Tetrahymena* Genome Database (118, 119), followed by analysis using the InterPro database (101). Conserved domains are highlighted, with colour indications described in the corresponding key; Conserved Walker A and Walker B motifs in MSH4 and MSH5 are underlined, MLH3 conserved metal-binding motif is underlined.

Appendix G
Conjugation pairing efficiencies of WT and MMR KO strains