**MSH3** promotes dynamic behavior of trinucleotide repeat tracts *in vivo*

Gregory M. Williams¹ and Jennifer A. Surtees¹,²*

¹Department of Biochemistry, School of Medicine and Biomedical Sciences, State University of New York at Buffalo, Buffalo, NY 14214

²Genetics, Genomics and Bioinformatics Program, School of Medicine and Biomedical Sciences, State University of New York at Buffalo, Buffalo, NY 14214

*Corresponding author:
Jennifer A. Surtees
Department of Biochemistry
University at Buffalo, State University of New York
619 Biomedical Research Building
3435 Main Street
Buffalo, NY 14214
Phone: (716) 829-6083; Fax: (716) 829-2725; Email: jsurtees@buffalo.edu
Abstract

Trinucleotide repeat (TNR) expansions are the underlying cause of more than forty neurodegenerative and neuromuscular diseases, including myotonic dystrophy and Huntington’s disease, yet the pathway to expansion remains poorly understood. An important step in expansion is the shift from a stable TNR sequence to an unstable, expanding tract, which is thought to occur once a TNR attains a threshold-length. Modeling of human data has indicated that TNR tracts are increasingly likely to expand as they increase in size, and to do so in increments that are smaller than the repeat itself, but this has not been tested experimentally. Genetic work has implicated the mismatch repair factor MSH3 in promoting expansions. Using S. cerevisiae as a model for CAG and CTG tract dynamics, we examined individual threshold-length TNR tracts in vivo over time, in MSH3 and msh3Δ backgrounds. We demonstrated, for the first time, that these TNR tracts are highly dynamic. Furthermore, we established that once such a tract has expanded by even a few repeat units, it is significantly more likely to expand again. Finally, we show that threshold-length TNR sequences readily accumulate net incremental expansions over time, through a series of small expansion and contraction events. Importantly, the tracts were substantially stabilized in the msh3Δ background, with a bias toward contractions, indicating that Msh2-Msh3 plays an important role in shifting the expansion-contraction equilibrium towards expansion in the early stages of TNR tract expansion.
**Author Summary**

Understanding the dynamics of threshold-length trinucleotide repeat (TNR) sequences allows us to understand the critical transition from a short, stable repeat to an expanding repeat that is pathogenic. We demonstrated that a small increase in TNR length significantly increased its propensity to continue increasing; these increases occur in small increments. We also demonstrated that *MSH3* plays a significant role in promoting these small incremental expansions. Loss of *MSH3* slowed the rate of expansion, and concomitantly increased the probability of a contraction in the TNR sequence. These findings make *MSH3* an attractive therapeutic target in preventing TNR expansions.
Introduction

The expansion of trinucleotide repeat (TNR) sequences is the underlying cause of over forty neurodegenerative and neuromuscular diseases (CASTEL et al. 2010; McMURRAY 2010). TNR sequences made of (CNG)$_n$ repeats are of particular interest, because of their role in causing Huntington’s disease (HD) and myotonic dystrophy type 1 (DM1), as well as a number of other diseases (McMURRAY 2010). TNR tracts within the normal range (which is tract-dependent) are stably maintained within that range (Fig.1). However, through a mechanism(s) that remains unclear, a TNR tract can expand, increasing the number of repeats within the tract. Initially this brings the tract into a “threshold-length” range (GANNON et al. 2012; CONCANNON and LAHUE 2014) (Fig. 1), in which these somewhat longer tracts are not pathogenic, but are increasingly susceptible to expansion; these individuals are carriers for disease. Once a tract has expanded sufficiently, it crosses a threshold; tracts above this threshold (which is disease-specific) are pathogenic and cause disease (Fig. 1). As the size of the tract increases, it becomes increasingly unstable and prone to changes in length, particularly expansions.

The dynamic behavior of TNR tracts that are within the threshold range (i.e. more susceptible to expansion) and the manner in which they occur in vivo remain unclear. Studies of TNR tract length changes have largely relied on end point experiments and therefore do not address dynamic behavior of the tracts, i.e. the rate at which tracts continue to expand. Modeling of human data has predicted that threshold-length TNR tracts will continue to increase in length, in increments smaller than the repeat itself (HIGHAM et al. 2012; MORALES et al. 2012; HIGHAM and MONCKTON 2013), although this has never been demonstrated directly. Single sperm typing studies in humans demonstrated small expansion and contraction events (1-2 repeats) in TNR sequences in Kennedy’s disease, Huntington’s disease and myotonic dystrophy patients (ZHANG et al. 1994; LEEFLANG et al. 1995; LEEFLANG et al. 1999; MARTORELL et al. 2004), particularly in alleles near the pathogenic threshold (ZHANG et al. 1994; LEEFLANG et al. 1995; CASTEL et al. 2010). These observations are consistent with the existence of an
equilibrium between expansion and contraction events. As the tract length increased in sperm cells, there was significant bias toward expansions versus contractions. The sizes of the observed expansion and contractions events were similar to our in vitro observations (1-2 repeats) (KANTARTZIS et al. 2012), in contrast to larger expansion events observed in non-dividing cells or post-mitotic neurons (McMURRAY 2010).

One factor known to contribute to (CNG)_n tract expansions is the mismatch repair (MMR) complex Msh2-Msh3 (MutSβ in mammals). Typically Msh2-Msh3 recognizes and binds insertion deletion loops (IDLs) that result from DNA polymerase slippage events, often within repetitive sequences (LOVETT 2004; LI 2008). These are then targeted for excision and re-synthesis. Strikingly, rather than correcting them, Msh2-Msh3 promotes TNR expansions in both mammalian somatic and germ cells (CASTEL et al. 2010; McMURRAY 2010). This difference is likely related to the propensity of (CNG)_n sequences to form secondary structures once they have slipped and are single-stranded, due to the inherent complementarity of the C’s and G’s within the repeat sequence (CASTEL et al. 2010; McMURRAY 2010), and the manner in which Msh2-Msh3 interacts with these unique structures (LANG et al. 2011). As the tract length increases, the potential complexity of the secondary structure increases (GACY and McMURRAY 1998; PEARSON and SINDEN 1998; SLEAN et al. 2013). Nonetheless, loss of MSH2 or MSH3 leads to a significant decrease in expansion events in mouse models of Huntington’s disease (MANLEY et al. 1999; OWEN et al. 2005) and myotonic dystrophy (VAN DEN BROEK et al. 2002; FOIRY et al. 2006). Similarly, Msh3 promotes expansions in human cells (GANNON et al. 2012; HALABI et al. 2012). We recently demonstrated a significant decrease in expansion of both CAG and CTG repeat tracts in Saccharomyces cerevisiae in an msh3Δ background (KANTARTZIS et al. 2012).

The role that Msh2-Msh3 plays in promoting TNR expansions remains unclear. Our in vitro results indicated that yeast Msh2-Msh3 interferes with proper Okazaki fragment processing by Rad27 (Fen1) and Cdc9 (Lig1) in the presence of a dynamic CTG or CAG repeat tract, leading to small incremental expansions, providing mechanistic insight into the
role of Msh2-Msh3 in replication-dependent TNR expansions (KANTARTZIS et al. 2012). This model predicts that Msh2-Msh3 modulates TNR dynamics \textit{in vivo}. Subsequently, Stevens et al. (STEVENS et al. 2013) demonstrated that Msh2-Msh3 can promote expansions \textit{in vitro} using human cell culture extracts, in a replication-independent manner, consistent with a role for Msh2-Msh3 in promoting expansions in somatic as well as germline cells (CASTEL et al. 2010; McMURRAY 2010).

To address the question of threshold-length tract behavior and the role that \textit{MSH3} plays in that behavior, we have established \textit{S. cerevisiae} as a model to visualize and characterize individual tract dynamics in dividing cells as a function of time, through multiple generations. We observed highly dynamic DNA sequences with progressive, incremental expansion in a \textit{MSH3} background. The TNR tract dynamics were substantially altered in the absence of \textit{MSH3}, resulting in more stable repeat sequences. These observations indicate that \textit{MSH3} is an important factor in promoting the transition of a TNR tract into the pathogenic range.

\textbf{Methods and Materials}

\textbf{Strains and Media}

All yeast transformations were performed using the lithium acetate method (GIETZ et al. 1992). Yeast strains were derived from the S288c background and are detailed in Table S1. In addition to the \textit{msh3Δ::KANMX} strain, constructed by amplifying a chromosomal \textit{msh3Δ::KANMX} fragment from the yeast deletion collection that was integrated into the \textit{MSH3} chromosomal location in FY86 (KANTARTZIS et al. 2012), we constructed a second \textit{msh3Δ} strain in FY86. The \textit{msh3Δ::hisG} strain was constructed by integration of a \textit{hisG-URA3-hisG} cassette into the \textit{MSH3} locus using pEAI88 (LYNDAKER et al. 2008). Integration was selected on synthetic media lacking uracil. The \textit{URA3} gene was selected against by growth on 5-FOA, which selects for recombination between the two \textit{hisG} moieties (ALANI et al. 1987), resulting in a single \textit{hisG} interrupting the \textit{MSH3} open reading frame. The tract dynamics in the two \textit{msh3Δ} backgrounds were indistinguishable.
TNR substrates were integrated into \textit{MSH3} and \textit{msh3}\textsuperscript{\textDelta} strains, as described previously (\textit{Miret et al.} 1998; \textit{Dixon et al.} 2004; \textit{Kantartzis et al.} 2012). Briefly, each plasmid was digested with \textit{Bsu36I} and then transformed to allow integration by homologous recombination at the \textit{LYS2} chromosomal locus. Integrants were selected on synthetic media lacking histidine. Single integration of the TNR substrate at the correct location was confirmed by Southern blot, as described previously (\textit{Dixon et al.} 2004). Three independent, single integrants for each substrate were selected for further analysis. This was repeated in all three strain backgrounds. Prior to performing the expansion assay, each strain was struck out on synthetic media lacking histidine and uracil to obtain single colonies. Individual colonies were analyzed by colony PCR to ensure that they contained the proper number of repeats (\textit{Dixon et al.} 2004). A subset was also sequenced to confirm the tract, using \textit{SO299} as a primer (5’ \texttt{ACTTGGGGAGAGGTGCG}) (\textit{Dixon et al.} 2004; \textit{Kantartzis et al.} 2012).

**Growth curves**

Saturated cultures were diluted 1:100 in synthetic media lacking histidine. Absorbance (\textit{A}\textsubscript{600}) of the cultures was measured every 45 minutes for 10 hours. The cultures were allowed to saturate overnight and were then diluted again and the growth curve repeated. There was no difference in the growth curves of the different strains tested or between the first and second days of growth, with a doubling time of approximately 140 minutes once the cells reached exponential phase (\textbf{Fig. S1} and data not shown). In 24 hours, the cells underwent approximately 6 cell divisions. For microcolony growth curves, individual cells were tracked using dissection microscope with a micromanipulator, with examination every 30 minutes for 10 hours. The daughter cells were removed from the mother following each cell division and placed elsewhere on the plate. There was no observable difference in the timing of the cell cycle between the \textit{MSH3} and the \textit{msh3}\textsuperscript{\textDelta} strains, which both exhibited a doubling time of approximately 140 minutes. Generation times of growth in liquid and on solid media were...
confirmed by counting the number of cells in culture and in colonies of different sizes, using a hemacytometer.

**Liquid Time Course Protocol**

Expanded CAG or CTG tracts in either the wild-type or msh3Δ background were selected by growth on synthetic complete (SC) media lacking histidine (to maintain the tract construct) and containing 5-FOA (to select for expansions that abrogate URA3 expression; US Biologicals), as described previously (MIRET et al. 1998; KANTARTZIS et al. 2012). The size of the tract was analyzed by colony PCR using SO295 and SO296 (KANTARTZIS et al. 2012). The PCR cycle consisted of 1 round at 95°C for 5 minutes, then 35 cycles of 95°C for 2 minutes, 53°C for 1 minute and 72°C for 1 minute, followed by a final extension at 72°C for 10 minutes. The PCR products were digested with SphI and AflII and then electrophoresed through a 12% polyacrylamide gel in 1X TBE buffer and the gel was stained with ethidium bromide (EtBr)(0.5 µg/ml) to determine tract size. The gels were imaged on a GelDoc system (BioRad) and quantified by ImageQuant (Molecular Dynamics).

Three independent colonies were selected for expansion events in each genetic background i.e. (CTG)_{25} and (CAG)_{25} tracts in MSH3 and msh3Δ. Each colony was resuspended in 106mL synthetic complete media lacking histidine (SC-his) and grown to saturation at 30°C with shaking. 5 ml of each saturated culture was used to prepare genomic DNA (gDNA). From each saturated culture, a logarithmically growing (log phase) culture and a stationary phase culture were established. For the logarithmic culture, 10 ml of SC-his was inoculated with 200µl of the saturated culture and grown in a 30°C shaker for 24 hours. This dilution was repeated every 24 hours for the duration of the time course. The remainder of each saturated culture was collected by centrifugation, the pellet was washed twice with sterile water then resuspended in 100mL of sterile distilled water. These stationary phase cultures were maintained in the 30°C shaker for the duration of the time course. The stationary phase cell pellet was washed and resuspended in fresh water every 48 hours.
Every 24 hours, 5 ml from each log and stationary phase culture was used to prepare gDNA. On Days 1, 7 and 14, dilutions from each culture were plated on SC-his to get individual colonies. PCR was performed, as described above, to amplify the tract using the gDNA from each time point as a template. Colony PCR was performed on individual colonies from Days 7 and 14.

**PCR controls**

PCR has been used extensively to examine TNR tract sizes (Zhang et al. 1994; Leeflang et al. 1995; Miret et al. 1998; Leeflang et al. 1999; Dixon et al. 2004). Nonetheless, we performed several controls to ensure that these observed expansions were not PCR artefacts. First, we performed PCR reactions using the CAG or CTG tract plasmid (used to integrate the tract into the yeast chromosome (Miret et al. 1998; Kantartzis et al. 2012)) as a template alongside every set of PCR reactions. A subset of these reactions is shown in Fig. S2 and clearly demonstrate the ability of Taq polymerase to accurately and reproducibly amplify a tract of 25 repeats. In over a hundred PCR reactions using the plasmid as a template, the major product was a 75 base pair fragment in every case. Minor products were occasionally observed at larger sizes and likely represent background due to polymerase error. Based on Southern blots of 20 independent plasmid amplification reactions, ≥ 85% (±1.5% S.E.M.) of the total product is the 75 base pair, unexpanded tract (Fig. S3, right side and data not shown). As an additional control, we performed five independent colony PCR reactions simultaneously, from the same colony, to assess reproducibility of Taq in amplifying both expanded and unexpanded tracts by colony PCR (Fig. S4). The resulting pattern of PCR products was very consistent and reproducible.

**Mixing experiment**

To determine whether the ratio of PCR products was representative of the input DNA template, we performed PCR reactions that contained different ratios of genomic DNA from cells with
either an unexpanded and expanded TNR tract. The DNA was amplified and then quantified (Fig. S5). Genomic DNA was prepared from cells carrying an unexpanded tract and from cells carrying an expanded tract, using standard techniques. The DNA concentration of each preparation was measured by NanoDrop. The genomic DNA from both backgrounds was used as a template for PCR at different concentration ratios. Ratios of 1:1, 1:5, 1:10 and 1:20 unexpanded:expanded and expanded:unexpanded were tested as templates in the standard PCR reaction, described above.

The ratio of unexpanded product:expanded product correlated with the input DNA, although at a 1:1 ratio there was a notable bias toward amplification of the unexpanded tract. This reassured us that we were not preferentially amplifying the expanded products and thereby overestimating the expansion rate. Furthermore, the PCR output was a reasonable representation of the input template and we are more likely to err on the side of underestimating expansions events when performing PCR with a non-homogenous template population. We do note that this may also lead us to underestimate expansions in the msh3Δ background as well.

**Colony PCR Time Course Protocol**

Expanded TNR tracts were selected and the tract size was analyzed and measured, as described above. Tracts of different length were selected for this experiment. Colonies with expanded tracts were incubated continuously at 30°C for 10-14 days. Colony PCR was performed every 24 hours and the tract analyzed by PCR and digestion. Colony PCR was performed either from the same side of the colony or from around the perimeter of the colony, with similar results. When taken from around the colony, the samples were taken from 4 locations in rotations, corresponding to 12 o'clock, 3 o'clock, 6 o'clock and 9 o'clock. Because the cells were always sampled from the perimeter of the colony, we assumed a generation time of ~140 minutes, as described above.
Microcolony PCR Time Course Protocol

Expanded TNR tracts were selected and the tract size was analyzed and measured, as described above. Colonies with expanded tracts were resuspended in sterile water. Each suspension was run in a line on an SC-his plate. Using a micromanipulator on a dissection microscope, three single cells from each suspension were selected for the time course. Each plate was incubated at 30°C for 15-20 hours, until the microcolony contained between 250-1000 cells; this was the size necessary for reproducible PCR results. Three individual cells were removed from the microcolony to continue the time course. (Three cells were selected to ensure that at least one microcolony arose in the next round.) The remaining cells were used for colony PCR, to assess the size of the TNR tract. The plate was then incubated at 30°C for another 15-20 hours, to allow growth of the four individual cells to a microcolony. One of the four was selected to propagate and continue the time course. The time course was continued through 14 cycles. At each step, tract sizes were analyzed by PCR, as described above. Time points in which no PCR amplification was observed were excluded from our analysis.

Southern blotting

To verify that PCR products contained the TNR tract, Southern blots were performed (SAMBROOK and RUSSELL 2001). Briefly, PCR reactions were electrophoresed through a 12% acrylamide gel (1X TBE). The gel was transferred to nylon membrane, cross-linked and hybridized with a radiolabeled probe, as described previously (DIXON et al. 2004; KANTARTZIS et al. 2012). The probe was a PCR product amplified from pBL69 (MIRET et al. 1998) with SO295 and SO296 and then digested with Sppl. The resulting fragment was radiolabeled using the Takara random priming kit. The membrane was washed, dried and exposed to a PhosphorImager screen and imaged and quantified, where applicable, in ImageQuant.

Mutation rate calculations
To calculate the mutation rates, the probability of a change in tract size was treated as a binomial distribution with $p =$ the proportion of tracts with a change in length and $q =$ the proportion of tracts with no length change. The rate of tract change was defined as the number of changes observed/ the number of cells examined per generation. For the liquid time courses, the mutation rate was derived from the number of tract changes observed per colonies tested divided by 126 generations per culture was used. For the colony and microcolony time courses, the proportion of tract changes was calculated as number of tract changes divided by the number time points examined divided by the number of generations from one time point to the next. Growth curve experiments in both liquid media and on plates were used to determine the number of generations for these calculations (Fig. S1 and data not shown). To calculate 95% confidence intervals (95% C.I.) on the mutation rates, the F statistic was used (ZAR 1999; FOSTER et al. 2006).

Plasmid retention assays
Three independent $MSH3$ isolates with large tracts, ranging from 33-48 repeats, were transformed with a low copy ($ARS CEN$) plasmid carrying the $KANMX$ cassette, which confers resistance to G418 (WACH et al. 1994). Three independent $MSH3$ isolates with small tracts, ranging from 25-30 repeats, were transformed with a low copy ($ARS CEN$) plasmid carrying the $NATMX$ cassette, which confers resistance to nourseothricin (GOLDSTEIN and MCCUSKER 1999). The two plasmids had indistinguishable retention rates in the absence of selection (data not shown). In each of three cultures, one isolate with a large tract and one isolate with a small tract were mixed 1:1 and were taken through a time course in SC-his for 6 days (see liquid time course protocol above), in the absence of selection for either plasmid. In the first set, a 40 repeat tract and a 25 (unexpanded) repeat tract were mixed. In the second set, a 38 repeat tract and a 31 repeat tract were mixed. In the third set, a 33 repeat tract and a 25 repeat tract were mixed. Every 24 hours, a sample of the culture was plated on SC-his media. Each individual colony was tested for growth on YPD media containing either G418 or
nourseothricin. The percentage of cells retaining resistance for one drug or the other is plotted in Fig. S6.

**Results**

The major goal of this study was to determine how a TNR tract behaves as it approaches a pathogenic length. To do this, we adapted the *in vivo*, *URA3*-based reporter system in *S. cerevisiae* that we used previously (MIRET *et al.* 1998; KANTartzis *et al.* 2012) and examined larger, threshold-length (CNG)$_n$ tract dynamics over time, in the presence and absence of *MSH3*. Previous work indicated that 25 repeat tracts approach the threshold-length in this system (MIRET *et al.* 1998; CONCANNON and LAHUE 2014). Therefore we started with (CTG)$_{25}$ and (CAG)$_{25}$ TNR tracts and selected single colonies with expanded tracts, as previously described (see Methods)(MIRET *et al.* 1998; KANTartzis *et al.* 2012). We took this analysis further by determining: 1) whether expanded tracts continue to expand, 2) at what rate tracts expand and 3) in what increments do expansions occur. Throughout this work we will follow the following convention: a CTG or CAG tract refers to the sequence of the nascent DNA on the lagging strand. Thus a CTG tract refers to CAG on the template strand and the newly replicated DNA will contain the CTG tract.

We took three different approaches to answering these questions, using PCR techniques to monitor tract length. First, we performed liquid time course experiments to determine whether additional expansions accumulated in replicating cells within a reasonable time frame (Figs. 3-5). Once a tract has expanded it is not possible to select for further expansions; this tested the feasibility of our approach. Second, we sought to observe dynamic tract size changes in a single colony over time (Figs. 6-9). The expanded tract of an individual, growing colony was monitored through multiple generations in the absence of selection. Third, we isolated individual cells from colonies with expanded tracts and monitored tract dynamics over a well-defined number of generations (Figs. 10, 11). In each case, approximate rates of tract changes were determined (Table 1). Combined, these three
approaches allowed us to monitor tract length dynamics through multiple generations, to document the type of tract changes that occurred \textit{in vivo} and the rates at which they occurred. Each set of experiments was done in the presence and absence of $\textit{MSH3}$, to determine its role in tract dynamic behavior.

Prior to selecting for expansion events, colonies are screened for unexpanded tracts by colony PCR (MIRET \textit{et al.} 1998; DIXON \textit{et al.} 2004; KANTARTZIS \textit{et al.} 2012). In the process of screening the tracts, we noted differences between the $\textit{MSH3}$ and $\textit{msh3}\Delta$ strain backgrounds. The ratio of unexpanded, expanded and contracted tracts was significantly different in the $\textit{MSH3}$ versus the $\textit{msh3}\Delta$ strains ($X^2$: $p$ value = 0.0038). In particular, the relative ratios of expanded:unexpanded tracts and expanded:contracted tracts were significantly higher in $\textit{MSH3}$ versus $\textit{msh3}\Delta$ strains (Fisher’s exact test: $p$ values = 0.0073 and 0.0061, respectively), consistent with a role for $\textit{MSH3}$ in promoting TNR expansions. These data also indicated that $\textit{msh3}\Delta$ strains exhibit an increase in contraction events, as previously suggested (SCHWEITZER and LIVINGSTON 1997; CASTEL \textit{et al.} 2010).

Distinct expansion tracts in CAG and CTG repeat tracts

To measure the size of each “initial expansion event” in CTG and CAG tracts, the TNR tract from a 5-FOA-resistant colony was amplified by PCR (MIRET \textit{et al.} 1998; DIXON \textit{et al.} 2004; KANTARTZIS \textit{et al.} 2012) and the product was digested with $\textit{SphI}$ and $\textit{AflII}$ (see Materials and Methods), generating a DNA fragment that contains only the TNR tract. Although PCR has been used extensively to examine TNR tract sizes (ZHANG \textit{et al.} 1994; LEEFLANG \textit{et al.} 1995; MIRET \textit{et al.} 1998; LEEFLANG \textit{et al.} 1999; DIXON \textit{et al.} 2004), we performed several controls to demonstrate that the observed expansions were not PCR artifacts (see Materials and Methods, Figs. S2-S5). With an unexpanded tract, the PCR product is 75 base pairs (25 CNG repeats); an expanded tract is larger and a contracted tract is smaller (Fig. S3). The size of each fragment was interpolated from a standard curve generated by the migration of low
molecular weight DNA standards (New England Biolabs) electrophoresed alongside the PCR products.

In *MSH3* cells, the median size of the initial expansion event, i.e. the expansion observed following selection on 5-FOA medium, was 42 repeats for the *CAG* tract (with a range of 31-67 repeats) and 33 repeats for the *CTG* tract (with a range of 26-43 repeats) (Fig. 2; Table S2), an increase of 17 and 8 repeats, respectively. The larger expansion size in the presence of the *CAG* tract is consistent with previous work (MIRET *et al.* 1998) and is likely due to the relative stability of *CTG* and *CAG* secondary structures. *CAG* repeats form less stable structures than *CTG* repeats (MARQUIS GACY *et al.* 1995; PEARSON and SINDEN 1996; BACOLLA *et al.* 2008) and therefore a larger expansion is predicted to be required to stabilize the secondary structure sufficiently to promote an expansion (MIRET *et al.* 1998). Importantly, these size ranges put the tracts near the pathogenic threshold for *CAG* and *CTG* repeats in coding regions in humans (Fig. 1) (LEEFLANG *et al.* 1999; CASTEL *et al.* 2010; MCMURRAY 2010).

In two different *msh3Δ* backgrounds (see Materials and Methods, Table S1), expansion events were less frequent, as previously reported (KANTARTZIS *et al.* 2012). The median initial tract length was 31 repeats in the *CAG* tract (with a range of 31-48 repeats) and 33.5 in the *CTG* tract (with a range of 28-39 repeats), an increase of 6 and 8 repeats, respectively (Fig. 2; Table S2). It is unclear why the size of the *CAG* tract initial expansion is significantly smaller in the *msh3Δ* versus the *MSH3* background (Fig. 2), while the *CTG* tract expansions were comparable in both backgrounds.

**Expanded TNR tracts continue to expand in vivo**

Using the same yeast reporter system, Rolfsmeier *et al.* demonstrated increased rates of *CTG* tract expansions and contractions as tract length increased (ROLFSMEIER *et al.* 2001). Therefore, we predicted that the initial expansion events that we observed, all ≥ 29 repeats, would similarly be less stable than the 25 repeat, increasing the probability of observing
additional expansion events over time. However, although we can select for the initial expansion event, it is not possible to select for additional expansion events in this system. Therefore, to test the feasibility of screening expanded CAG and CTG tracts for additional expansions, we performed fourteen day time course experiments in parallel liquid cultures (Fig. 3). Three independent isolates (single colonies) with expanded CAG or CTG repeat tracts in MSH3 or msh3Δ backgrounds, representing a single initial expansion event (Day 0 in Figs. 4, 5), were selected and used to inoculate starting cultures for a total of 12 independent time course experiments. Isolates with different initial tract sizes were selected for these experiments (Table S2). The growth curves of MSH3 and msh3Δ strains carrying either CAG or CTG repeat tracts were indistinguishable (Fig. S1 and data not shown); the cells went through approximately 6 generations every 24 hours. From each starter culture, we established parallel logarithmic (log) and stationary phase cultures, in liquid media (Fig. 3).

Genomic DNA was isolated from each log and stationary phase culture every 24 hours and used as a template for PCR to assess TNR tract sizes within the population of the culture. A representative time course is shown in Fig. 4A, top panel. Cells from each culture were also plated at days 7 and 14; individual colonies were used as a template for colony PCR to amplify individual TNR tracts at different time points (Fig. 4B, middle panel). There was no selection for expansion events during the time course, i.e. no 5-FOA. Therefore the tracts were free to remain stable, to expand or to contract.

In the MSH3 background, the cells readily accumulated additional expansions in actively dividing cells (Fig. 4); tracts in stationary cultures were substantially more stable (see below). In both CAG and CTG tract time courses, analysis of the genomic DNA revealed that the tract size exhibited a general shift in the population to a larger size (Fig. 4A, top). The band representing the major initial expansion event of 31 repeats (asterisk in Fig. 4A) became less prominent over the time course. At the same time, a higher molecular weight product accumulated and became more prominent as the time course progressed. We note that the larger fragment was present at the beginning of the time course, presumably due to additional
expansion events that occurred during the three days required to bring the cultures to saturation prior to setting up parallel cultures (see Materials and Methods, Fig. 3). Two lines of evidence suggest that this pattern is not a result of a selective advantage of cells with larger repeat tracts. First, there were only about 6 generations from one time point to the next (Fig. S1). Second, we performed plasmid retention assays in which cells with a short tract were transformed with a plasmid conferring resistance to ClonNAT and cells containing a long tract were transformed with a plasmid conferring resistance to G418. Equal numbers of each cell type were mixed and allowed to grow in the absence of selection for 6 days. Both plasmids were present at a 1:1 ratio throughout the time course (Fig. S6), indicating that neither strain has a selective advantage.

To determine whether the higher molecular weight tract represented 1) a single larger expansion event that accumulates within the population over time or 2) an average of multiple different-sized expansion events within the population, we performed colony PCR on individual colonies plated on days 7 and 14. Colony PCR clearly demonstrated that a variety of tract lengths existed within the population (Fig. 4A, middle) indicating multiple independent expansion events. CTG tracts from individual colonies revealed 43% and 47% of the CTG tracts had sustained an additional expansion event at days 7 and 14 of the time course, respectively. Similarly, the CAG tracts exhibited a high frequency of additional expansion; 42% and 78% expansion at days 7 and 14, respectively (Fig. 4A, lower). These frequencies and the growth rate, allowed us to calculate expansion rates of $1.2 \times 10^{-3}$ per cell generation (95% C.I., $9.0 \times 10^{-4}$ to $1.6 \times 10^{-3}$) for MSH3(CTG) and $3.1 \times 10^{-3}$ (95% C.I., $2.4 \times 10^{-3}$ to $3.4 \times 10^{-3}$) per cell generation in the MSH3(CAG) background (Table 1). These rates are ~2 orders of magnitude higher than we observed when we performed time courses starting with unexpanded (CTG)$_{25}$ or (CAG)$_{25}$ tracts (Table 1, Fig. S7). Notably, the rates of initial expansion (i.e. starting with an unexpanded tract) in the time course assays performed here were consistent with our previous measurements for the initial expansions using the selective assay (KANTARTZIS et al. 2012). These data indicate that small increases in tract length
significantly increase the rate of expansion. Furthermore, these rates may be
underestimates; the further expanded tracts that we observed could have arisen through more
than one expansion event. It is noteworthy that the CAG tract, which has a lower expansion
rate when starting at 25 repeats (Miret et al. 1998; Kantartzis et al. 2012), was more likely to
continue expanding than the CTG tract, perhaps a result of the longer starting tract sizes,
consistent with the idea that stability of the secondary structure that forms within the tract
affects the probability of expansion (Miret et al. 1998; Rolfsmeier et al. 2001). Once the
CAG tract has increased sufficiently to generate more stable secondary structure, it expands at
least as rapidly as the CTG repeat tract.

**Expanded tracts are more stable in msh3Δ background**

In contrast to the MSH3 strains, both CAG and CTG expanded tracts in the msh3Δ strains
were much more stable when propagated in liquid culture (**Fig. 4B**). We observed no
additional expansions in the msh3Δ background when the genomic DNA was analyzed (**Fig.
4B, top** and data not shown). Intriguingly, the analysis of the genomic DNA from one msh3Δ
time course revealed contraction events starting at Day 12 (**Fig. 4B, top**). Colony PCR of
tracts derived from individual cells similarly revealed very few additional expansion events
(12% for CTG tracts, 5% for CAG tracts) (**Fig. 4B, middle, lower**); the CTG tracts were as
likely to contract as to expand (12% expansions versus 8% contractions). The expansion rate
for msh3Δ(CTG) was $3.1 \times 10^{-4}$ per cell generation (95% C.I., $1.3 \times 10^{-4}$ to $6.0 \times 10^{-4}$), 4-fold
lower that in the MSH3 background. Similarly, the expansion rate in the msh3Δ(CAG)
background was ~20-fold lower than in MSH3 strains, at $1.3 \times 10^{-4}$ (95% C.I., $2.9 \times 10^{-5}$ to $3.7 \times
10^{-4}$) per cell generation (**Table 1**).

These data demonstrate that CNG repeat tracts become increasingly unstable and
much more susceptible to expansion once it has crossed the “stability” threshold, i.e. ≥29
repeats in this system, particularly in the presence of MSH3. These data also indicate that the
presence of MSH3 promotes not only the initial expansion event (or events) that allow for
selection in this system, but also subsequent expansion events, at least within this threshold range. Therefore we can use expanded tracts as a starting point to examine tract dynamics; there are likely to be changes within a time frame of several days. Furthermore, the initial expansions are relatively small, increasing to a total of $\leq 50$ repeats, representing early events in the progression of an expanded tract.

**Expansions were enhanced in dividing cells**

In contrast to CNG repeat tracts in replicating cells, the tracts from the stationary phase cultures in both the *MSH3* and *msh3Δ* backgrounds were more stable ([Fig. 5](#)), indicating that a significant proportion of the expansion events are dependent on DNA replication. Based on analysis of the genomic DNA from *MSH3* cultures, the major initial expansion event (asterisk) remains the major PCR product throughout the time course; there is little or no accumulation of additional expansion events ([Fig. 5A, top](#)). Colony PCR from individual colonies similarly revealed that the tracts were very stable in stationary phase ([Fig. 5A, middle, lower](#)). We observed that 7% and 8% of the CTG tracts were further expanded at days 7 and 14, respectively, while 12% and 15% of the CAG tracts had additional expansions at days 7 and 14, respectively.

The frequency of expansions in the *msh3Δ* background was also lower in stationary phase cultures ([Fig. 5B](#)). There were neither expansions nor contractions when the genomic DNA was analyzed. And there were very few additional expansions observed from the colony PCR analysis, again confirming that DNA replication is important in promoting these additional expansion events. We do note, however, that even in stationary phase, there were more additional expansion events in the *MSH3* background than in the *msh3Δ* background. With the CTG tract, we observed 8% and 1% expansions at days 7 and 14, respectively; we observed no expansions in the CAG stationary phase cultures ([Fig. 5B, lower](#)). This is consistent with the observation that Msh2-Msh3 promotes CNG expansions in post-mitotic (non-dividing) tissues in mammalian systems (McMurray 2010; Kovalenko *et al.* 2012).
Expanded TNR tracts increase in discrete steps

To follow tract dynamics as a function of time, we selected independent expanded tracts on 5-FOA medium and confirmed their sizes by PCR, as described above. Colonies with a range of initial tract sizes (i.e. initial expansion events) were selected for further analysis. We chose to focus on CTG tracts because the range of initial expansion events was more similar in the MSH3 and msh3Δ backgrounds than with CAG tracts (Fig. 2, Table S2). Each colony was then transferred to non-selective media (no 5-FOA) so that the tract was free to expand or contract (Fig. 6). Colony PCR was performed on the same colony every 24 hours for 10-14 days to assess tract length. Additional expansions were observed in the majority of the time courses and the increase in the size of the tract was progressive (Figs. 7A, 8A). While multiple tract sizes were often observed in the same time point, overall the size of the predominant tract increased over time, with a concomitant loss of the smaller-sized tracts. We refer to this as a progressive tract length change. Typically, 2-3 discrete progressive changes, or increments, were observed over 14 days, indicating that the expansions occur in steps. We confirmed that these PCR products contained the TNR tract by Southern blot, using a tract-specific (CTG) probe (Fig. 7A, right). We also performed PCR controls to rule out the possibility that larger tracts were preferentially amplified (see Materials and Methods, Fig. S5).

In Fig. 8A, we plotted the progressive changes in tract length as a function of time for all of the MSH3(CTG) time courses that we evaluated. There is a clear upward trend in tract length in the majority of the time courses, although contractions and stable tracts are also evident. Importantly, these changes were discrete increases in size, which we refer to as “steps”. Overall, the tracts appeared to expand in relatively small increments, or steps, over time (Figs. 7A, 8A). Even cells with very large initial expansions exhibited incremental increases in this range. For example, two colonies (Fig. 8A; B4 and C4) with initial expansions of 38 repeats (for total lengths of 63 repeats) had subsequent increases of only 8 and 9 repeat lengths, respectively. Conversely, one 70 repeat tract (Fig. 8A; D4) sustained a 24 repeat
increase. Therefore in our data set there does not seem to be a strict correlation between tract length and the size of additional expansion events in the tract length range that we investigated. All the increases were smaller than the initial tract length. These data therefore suggest that the longer tract expansions are the result of multiple expansion events.

To assess the size of each step and determine whether there is a correlation in step size with respect to starting tract length, we plotted the number of additional repeat units we observed at each tract length change for all the time courses plotted in Fig. 8A (Fig. 9A). Thus the pink circles demonstrate the number of repeat units we observed in each initial expansion event, i.e. after selection on 5-FOA. The median size of the initial expansion was 24 base pairs (8 repeats). An additional increase in tract length, i.e. the first ‘step’ beyond the initial expansion, was plotted using the orange squares and the subsequent increases, or second, third and fourth ‘steps’, were similarly plotted. Thus only the time courses that exhibited progressive tract length changes are represented in this figure. We note that there was not a strict correlation in the timing of the tract length changes (Fig. 8A) and therefore the size of the tract changes was plotted independent of time point. When all expansions were taken into account, the median size of the tract length increase, or step, was 24 base pairs (8 repeats); the average increase was 11 repeats (±18 repeats; 2 standard deviations) but the range was large: 2 to 47 repeats. The median size of each step was 8 repeats and did not increase with increasing tract size, within the range of tract lengths that we observed (Fig. 9A). In fact, the increase in tract size decreased after the initial expansion, as did the range in expansion events.

When individual msh3Δ colonies with an expanded TNR tract were followed over time, the TNR tracts appeared more stable (Figs. 7B, 8B, Fig. 9B), consistent with our observations in the population-based time course (Fig. 4). In 64 time courses, we observed no evidence of progressive expansion (Figs. 7B, 8B), as defined above with the concomitant loss of the smaller tract length. We did observe additional PCR products in some cases, but this resulted in a mixed population, with ≥ 75% of the tract remaining at the original expansion size (Fig. 9A).
The results of the mixing experiments (Fig. S5B) suggest that this ratio indicates an approximately 5-fold excess of unexpanded tract template in the reaction. In these events, we observed a total of 1 to 2 tract species in addition to the original tract size, which continued to dominate the population. We also observed contractions of the predominant PCR product in 7 independent time courses (Fig. 8B). In contrast, no progressive contractions were observed in the MSH3 background, although some individual contractions were observed (Fig. 8A). Southern blotting confirmed the presence of the CTG tract in each of these DNA fragments (Fig. 7B and data not shown). Overall, taking only progressive tract length changes into account, the rate of change in tract length in msh3Δ cells was 3.6 x 10⁻³ per generation (95% C.I.: 1.5 x 10⁻³ to 7.2 x 10⁻³); the changes observed were all contractions. In contrast, the rate of tract length change in MSH3 cells, again including only progressive tract length changes, was 1.5 x 10⁻² per generation (95% C.I.: 1.2 x 10⁻² to 1.9 x 10⁻²), 4-fold higher than in the msh3Δ cells, and the observed changes were almost exclusively expansions (Table 1). We note that these are likely underestimates because minor expansion and contraction products were observed in both MSH3 and msh3Δ time courses. Furthermore, the extent of expanded products observed may be dependent on the section of the colony sampled. Nonetheless, the presence of Msh2-Msh3 appears to shift the dynamic equilibrium toward expansion over time, ultimately resulting in longer and longer TNR tracts. The absence of the complex stabilizes the tract and permits an increase in contraction events, possibly as a result of the MMR deficiency (see Discussion).

**TNR tracts are highly dynamic as they expand**

The incremental expansions of the TNR tract observed within individual colonies were smaller than the tract length, but were typically larger than the increments we observed in vitro (KANTARTZIS et al. 2012), although they were on the same order of expansions previously observed by single sperm typing in humans (LEEFLANG et al. 1995; LEEFLANG et al. 1999). To determine whether each expansion event represented a single larger event or the sum of
multiple small events, indicating a highly dynamic TNR tract, we initiated time courses with a
single cell, rather than a potentially mixed population within a colony. First, we selected a
colony that contained an expanded TNR tract. From this colony, we isolated individual cells
using a micromanipulator (Fig.10). These single cells were allowed to go through
approximately 8-10 cell divisions (and therefore 8-10 rounds of replication), resulting in a
microcolony of ~250-1000 cells. Individual cells from each microcolony were selected to
propagate a new microcolony to continue the time course. The remainder of the microcolony
was used as a template for PCR.

In the MSH3 background, starting with tracts between 30 and 33 repeats, we observed
very dynamic behavior of the tract, with several small changes in tract size occurring over the
course of two weeks (Fig. 11A; Table 2). Multiple products were frequently observed at each
time point, indicating a mixed tract population within the microcolony. We quantified the
predominant tract at each time point and observed both expansions and contractions in similar
quantities (slight bias toward expansions) and of similar sizes, many of which were 1, 2 or 3
TNR repeats in length (i.e. 3, 6 or 9 nucleotides) (Table 4). Although the total number of
events is low, we calculated an approximate rate of tract size change per generation
(expansions and contractions) of $3.5 \times 10^{-2}$ per generation (95% C.I.: $2.8 \times 10^{-2}$ to $4.2 \times 10^{-2}$),
assuming 10 generations per time point (Table 1). The changes in tract size observed on this
time scale were smaller than those observed in the colony time course and more closely
resembled the size of expansions observed in vitro in the presence of Msh2-Msh3, Rad27
(Fen1) and Cdc9 (Lig1) (KANTARTZIS et al. 2012). Therefore, in a wild-type (MSH3)
background threshold-length TNR tracts demonstrated a tendency to gradually increase in
size. Notably, even in tracts where there was no net change in tract length at the end of the
time course, the tract length did change during the time course (i.e. C, D, E and G in Table 2).

In the absence of MSH3, the tracts appeared more stable overall (Fig. 11B; Table 3)
with an estimated mutation rate per generation of $1.6 \times 10^{-2}$ per generation (95% C.I.: $1.1 \times 10^{-2}$
to $2.2 \times 10^{-2}$) (Table 1), half the rate observed in the presence of MSH3. We do note that
additional PCR products were observed in some of these time courses. However these additional bands did not change in size over the course of the experiment, indicating stability. Furthermore, in several cases, including the lower panel of Fig. 11B, Southern blots indicated that the upper bands did not contain TNR sequence (data not shown). When changes did occur, the median size of the increments was similar to that observed in the MSH3 background. However, the tracts in the msh3Δ strains were more likely to contract than to expand (10 expansions:18 contractions) (Table 4), in contrast to roughly equal proportions of expansions and contractions in the MSH3 background, with a bias toward expansions (38 expansions:30 contractions) (Table 4).

Discussion

We have, for the first time in any system, documented the dynamic behavior of threshold-length TNR tracts and tracked them over time in dividing cells. The rate of change in the size of threshold-length TNR tracts is remarkable, at 2-3 orders of magnitude higher than the stable (CNG)$_{25}$ repeats in yeast (MIRET et al. 1998; ROLFSMEIER et al. 2001; KANTARTZIS et al. 2012). Thus small increases in tract length led to substantially higher expansion rates. The rapid rate of change in tract length, both expansions and contractions, was particularly striking when threshold-length tracts were examined in microcolonies derived from a single cell. In contrast, we did not observe similar changes in an unexpanded (CTG)$_{25}$ tract (Fig. S8).

Previous studies have proposed a model for TNR expansions in dividing cells that involved incremental expansions (LEEFLANG et al. 1995; LEEFLANG et al. 1999; MARTORELL et al. 2004; KANTARTZIS et al. 2012). Here we have shown directly that incremental expansions do in fact occur in vivo within CTG repeat tracts. The size of the increments varies, but can be as small as a single repeat unit. This is consistent with the behavior of CAG repeats at both the Huntington’s and androgen receptor loci in humans (ZHANG et al. 1994; LEEFLANG et al. 1995). Recent work modeling human DM1 expansion data has also demonstrated that these expansions are likely to occur through small incremental steps (HIGHAM et al. 2012; MORALES
et al. 2012; Higham and Monckton 2013). In induced pluripotent stem cells (iPSC) from HD and DM patients, Du et al (2013) observed that (CAG)\textsubscript{46} and (CTG)\textsubscript{57} tracts were stable after 12 and 16 passages, respectively (Du et al. 2013). For comparison, yeast cells went through approximately 120 cell divisions during the 2 week time course experiments in this study. However, a larger CTG tract (~126 repeats) expanded by ~ 1 repeat increments in each passage; a (CTG)\textsubscript{773} tract expanded in larger increments (Du et al. 2013). These data are consistent with a step-wise increase in the early stages of TNR tract expansion.

Importantly, the absence of MSH3 alters the dynamics of the tract in this model system. Overall, in actively replicating cells, both the CAG and CTG tracts are more stable in the msh3\textsubscript{Δ} backgrounds, consistent with previous observations (Owen et al. 2005; Kantartzis et al. 2012). Expansions did occur, but at a significantly lower rate than in the MSH3 background, as described previously (Kantartzis et al. 2012). Furthermore, msh3\textsubscript{Δ} cells were much less likely to incur an additional expansion and when additional expansions did occur, they were less likely to become the predominant population (Figs. 3-11), despite the fact that the cell division time for the two strains was indistinguishable. Therefore loss of MSH3 reduces the probability of incurring both an initiating and a secondary expansion event. On the other hand, contractions events were observed more frequently in msh3\textsubscript{Δ} strains (e.g. Fig. 8; Table 4). This is consistent with Schweitzer and Livingston (1997), who demonstrated an increase in TNR contractions in msh2\textsubscript{Δ} cells, using a different yeast reporter system (Schweitzer and Livingston 1997). The increased rate of contraction may be related to the loss of IDL MMR in the absence of MSH3; Msh2-Msh3 preferentially corrects deletion events, i.e. loops on the template strand, during DNA replication (Schweitzer and Livingston 1997; Sia et al. 1997; Romanova and Crouse 2013). These observations support the suggestion that Msh3 is a potential therapeutic target (Castel et al. 2010) and underscore the need to develop a better understanding of tract dynamics in different genetic backgrounds. It will be of interest to determine whether the size of the contractions vary with MSH3 status.
It is striking that, in the absence of selection for tracts larger than 30 repeats, we nonetheless observed a strong bias toward expansions in the \textit{MSH3} background, with very few contractions. This is distinct from previous observations in yeast, where there was a strong bias toward contractions (\textsc{Maurer et al.} 1996; \textsc{Miret et al.} 1997; \textsc{Rolfsmeier et al.} 2001), although when expansions were observed, they were relatively small, ranging from 2 to ~20 repeat lengths (\textsc{Maurer et al.} 1996; \textsc{Miret et al.} 1998), consistent with our observations. We performed a preliminary contraction assay, and observed rates of contraction similar to those previously observed (\textsc{Miret et al.} 1997), indicating that there is nothing inherent in our strain background that predisposes toward expansions (data not shown), but rather that the size of the tracts themselves led to the shift toward expansions. As such, the threshold-length tract behavior that we observed in yeast mimics what has been observed in human sperm cells and iPSCs and is therefore a good model for events in dividing cells. In this model we can look at large numbers of cells over many generations, making this a very flexible and informative approach to examining TNR tract dynamics. It is important to note that Msh2-Msh3-mediated expansions are just one mechanism by which TNR tracts increase in length. This yeast system is amenable to examining the contribution of additional genetic factors on tract dynamics, in both deletion and mutation backgrounds (e.g. (\textsc{Debacker et al.} 2012; \textsc{Concannon and Lahue} 2014)).

Although this study focused on actively replicating cells, we note that the frequency of additional expansions was also elevated in stationary phase cultures relative to the unexpanded tract (\textbf{Fig. 5A}); ~10-15\% of the tracts had an additional expansion event in the \textit{MSH3} background. Notably, this decreased in the absence of Msh3 (\textbf{Fig. 5B}). Msh2-Msh3 has been shown to be important in promoting somatic tract expansions in post-mitotic (non-replicating) neurons, specifically in medium-spiny striatal neurons (MSN) in a mouse model for HD (\textsc{Kovalenko et al.} 2012). Deletion of \textit{Msh2} eliminated the majority of \textit{CAG} expansions in the Huntington’s disease gene (\textit{HTT}) in the MSN. The decrease in expansions correlated with a reduction in \textit{HTT CAG}-dependent phenotypes in the mice. In another study, \textit{Msh3}
polymorphisms that altered Msh3 protein levels in mouse striatal tissues also altered the stability of the HD transgene CAG repeat tract: elevated Msh3 levels correlated with increased tract size (Tomé et al. 2013). Recent work has suggested that CAG expansions in somatic tissue may also occur in an incremental manner (Lee et al. 2011). A cohort of Huntington’s disease CAG knock-in mice was sampled at 2, 5, 9, 12 and 16 months of age and CAG tract length was examined in the liver and striatum. In these tissues, there was an average increase in tract size of 1 repeat per month, although there was a much broader distribution in the striatum than in the liver. In vitro, human Msh2-Msh3 (MutSβ) stimulated expansions in a replication-independent system (Stevens et al. 2013). Therefore monitoring TNR tracts in stationary phase yeast cultures may be a powerful way to model tract dynamics in non-proliferating cells and elucidate the mechanism(s) of expansions in somatic cells.

Although we have not tested the effect of msh2Δ on dynamic TNR tract behavior, our in vivo (this work; Kantartzis et al. 2012) and in vitro data (Kantartzis et al. 2012) examining TNR tracts are consistent with a model in which TNR secondary structures are bound and stabilized by Msh2-Msh3. Increased tract length increase the probability that alternative DNA structures will form, perhaps during Okazaki fragment processing or base excision repair, leading to the recruitment of Msh2-Msh3. So why are the structures not repaired? In vitro work with human MutSβ (Msh2-Msh3) has indicated that the Msh2-Msh3/TNR complex is distinct from the Msh2-Msh3/loop complex (Owen et al. 2009; Lang et al. 2011). Similarly, Msh2-Msh3 exhibits distinct nucleotide turnover behavior when bound to MMR versus double-strand DNA break repair (DSBR) substrates (Kumar et al. 2013; Kumar et al. 2014). It is not yet clear how these proposed kinetic differences could translate into increased expansions, but we favor the possibility that Msh2-Msh3-mediated ATP-binding and hydrolysis activities are altered in the presence of the TNR substrate, disrupting error-free MMR. Altered Msh2-Msh3 activity could block repair altogether or recruit an error-prone repair process that ultimately retains the insertion. Recent work has demonstrated that mammalian MutLγ (Mlh1-Mlh3) is required for somatic CAG expansions in an HD mouse; MutLα (Mlh1-Pms2) was dispensable
(PINTO et al. 2013). Intriguingly, Rogacheva et al demonstrated that yeast Mlh1-Mlh3 enhances Msh2-Msh3 DNA-binding activity, while Msh2-Msh3 stimulates Mlh1-Mlh3 endonuclease activity (ROGACHEVA et al. 2014). Therefore enhanced Msh2-Msh3 binding to TNR structures could inappropriately recruit Mlh1-Mlh3, during replication or outside of S phase, and stimulate its endonuclease activity, leading to gap formation and DNA repair synthesis that incorporates incremental expansions (GOMES-PEREIRA et al. 2004; PLUCIENNIK et al. 2013).

Disruption of Msh2-Msh3 has been proposed as an attractive therapeutic target for TNR expansion disease such as HD or DM1 (PANIGRAHI et al. 2005; PANIGRAHI et al. 2006; CASTEL et al. 2010). Loss of MSH3 leads to a decrease in expansion events and promotes contractions. However, disruption of Msh2-Msh3 will have a negative impact on overall genome stability; in addition to MMR, Msh2-Msh3 has been implicated in DSBR (SURTEES et al. 2004; PARK et al. 2013; VAN OERS et al. 2013) and interstrand DNA cross-link repair (TAKAHASHI et al.; BARBER et al. 2005; ZHAO et al. 2009). Notably, many neuronally expressed genes contain repetitive sequences in their regulatory regions (BACOLLA et al. 2008), the stability of which is likely at least partially dependent on Msh2-Msh3. Therefore, in considering Msh2-Msh3 as a molecular target, a clear mechanistic understanding of Msh2-Msh3 function in promoting repeats is vital to ensure that a balance is struck between promoting genome stability and genome instability in developing rational therapeutic strategies.
Acknowledgements

We are grateful to Drs. Mark Sutton, Robert Lahue, Catherine Freudenreich and Eric Alani for comments on this manuscript and to members of the Surtees lab for discussions of this work. Work in the Surtees lab is supported by NIH GM087459. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References


BACOLLA, A., J. E. LARSON, J. R. COLLINS, J. LI, A. MILOSAVLJEVIC et al., 2008 Abundance and length of simple repeats in vertebrate genomes are determined by their structural properties. Genome Research 18: 1545-1553.


HIGHAM, C. F., F. MORALES, C. A. COBBOLD, D. T. HAYDON and D. G. MONCKTON, 2012 High levels of somatic DNA diversity at the myotonic dystrophy type 1 locus are driven by ultra-frequent expansion and contraction mutations. Human Molecular Genetics 21: 2450-2463.


KUMAR, C., R. EICHMILLER, B. WANG, G. M. WILLIAMS, P. R. BIANCO et al., 2014 ATP binding and hydrolysis by Saccharomyces cerevisiae Msh2–Msh3 are differentially modulated by mismatch and double-strand break repair DNA substrates. DNA Repair 18: 18-30.


Figure Legends

**Figure 1. Size of trinucleotide repeat (TNR) tract determines disease phenotype.** The normal range of a TNR tract is present in unaffected individuals. This range varies depending on the TNR sequence and position with respect to the relevant gene. As the TNR tract gets larger, it enters the threshold or pre-mutation range. These individuals are typically unaffected by disease but are carriers because these somewhat larger tracts are prone to further expansion. Once the TNR tract expands into the affected range, individuals are symptomatic and are affected by disease.

**Figure 2. Distribution of tract size flowing initial expansion.** Whisker plot of the median tract size following selection on 5-FOA plates in *MSH3* and *msh3Δ* strains. The (CAG)\textsubscript{25} and (CTG)\textsubscript{25} tracts were examined. The plot was created using GraphPad Prism. Both the median and mean size of the CAG tracts are statistically different from the CTG tracts (Wilcoxon rank; t-test, \( p < 0.0001 \)). The median and mean CAG tract length in the *MSH3* background is statistically different from the CAG tract in the *msh3Δ* background, whereas the CTG tract size is similar in both strain backgrounds (Wilcoxon rank; t-test, \( p < 0.0001 \)).

**Figure 3. Schematic of the liquid time course experiment protocols.** Individual colonies with TNR tract expansions were selected on plates containing 5-FOA; the tract increase was confirmed by colony PCR in each case. A single colony with an expanded tract was used to inoculate a large culture that was grown to saturation over 72 hours (~18 generations). From this population, genomic DNA (gDNA) was isolated. Then parallel logarithmic and stationary phase cultures were established (see Materials and Methods) and propagated for 14 days. Genomic DNA was prepared from each culture every 24 hours (~6 generations) and then subjected to PCR to evaluate tract size. A sample from each logarithmic and stationary phase cultures was diluted and plated to isolate individual colonies (~26 generations) which were
then subjected to PCR to assess tract length. This experiment is akin to a mutation accumulation experiment, although we limited the number of cell divisions (~6/time point) in order to mitigate any fitness effects of tract changes. This is an endpoint experiment that indicated that additional expansions are observable in a 2 week time span. Further, it allowed us to observe general trends in the population as well as looking at individual tract lengths from single colonies. See Materials and Methods for additional details.

**Figure 4. Liquid time course experiments in MSH3 and msh3Δ logarithmic cultures to determine population tract dynamics.** TNR expansion events in the *MSH3* or *msh3Δ* background were selected and confirmed by PCR (see Materials and Methods) and logarithmic phase cultures were established from individual isolates (see Fig. 3 for cartoon).

For **A. MSH3(CTG)** and **B. msh3Δ(CTG): Top panel:** The TNR tract from genomic DNA isolated from the *MSH3* logarithmic phase cultures was amplified, digested and analyzed by electrophoresis. In these examples, the initial expansion (indicated by the asterisk) contained 31 repeats for *MSH3* and 37 repeats for *msh3Δ*. The numbers across the top of the gel indicate the day of the time course. The boxed region in **A** (*MSH3*) indicates the progressive accumulation of a larger tract. The boxed region in **B** (*msh3Δ*) indicates contraction events that result in a smaller tract. The cultures went through ~6 generations in a 24 hour period (see Materials and Methods). **Middle panel:** Samples from Day 14 cultures were plated on minimal media lacking histidine to obtain individual colonies. Colony PCR was performed to amplify the TNR tract from colonies to determine individual tract lengths within the population. The arrows indicate tracts that have incurred an additional expansion. The number below each tract indicates the number of repeats within each tract. Control PCR reactions using the TNR plasmid (C) were performed alongside each set as a marker for the 75 base pair (bp) tract (25 repeats). We occasionally observed higher molecular weight products with the control plasmid, but these represent a minor population of ≤ 15% of the total PCR product (**Figs. S2, S3**). Asterisk indicates the predominant expansion product in the starting culture.
**Bottom panel:** Summary of expansion frequency at Days 7 and 14 in the different genetic backgrounds tested, based on PCR amplification of tracts from individual colonies.

**Figure 5. Liquid time course experiments in MSH3 and msh3Δ stationary phase cultures to determine population tract dynamics.** Stationary phase cultures were maintained in parallel with logarithmic cultures (see Fig. 3 for cartoon). The examples shown here for A. MSH3 and B. msh3Δ match those shown in Fig. 3. **Top panel:** The TNR tract was amplified from genomic DNA isolated from the stationary phase culture, digested and analyzed by electrophoresis. In both MSH3 and msh3Δ, the population remained unchanged over time. The numbers across the top of the gel indicate the day of the time course. **Middle panel:** Samples from Day 14 cultures were plated on minimal media lacking histidine to obtain individual colonies. Colony PCR was performed to amplify individual TNR tracts from the stationary phase population to determine tract lengths. These tracts were quite stable. Control PCR reactions using the TNR plasmid (C) were performed alongside each set. Asterisk indicates the predominant expansion product in the starting culture. **Bottom panel:** Summary of expansion frequency at Days 7 and 14 in the different genetic backgrounds tested, based on PCR amplification of tracts from individual colonies. For more detail, see Table S4.

**Figure 6. Schematic of the colony time course experiment protocol.** Individual colonies with TNR tract expansions were selected on plates containing 5-FOA; the tract increase was confirmed by colony PCR in each case. A single colony with an expanded tract was transferred to non-selective media and allowed to continue to grow. For each time course, the same colony was subjected to colony PCR every 24 hours (~10 generations), sampling from the perimeter of the colony where the cells continued to grow. This approach allowed us to examine the dynamics of a single TNR tract as a function of time. Because the PCR is derived from a sub-sample of an actively dividing colony, we are necessarily examining a mixed
population albeit derived from the same original cell. See Materials and Methods for additional details.

**Figure 7. Progressive expansion events in MSH3 but not in msh3Δ.** A. **MSH3:** Expansion events were selected in the MSH3 \((CTG)_{25}\) background and confirmed by PCR. Individual colonies were followed over several days; colony PCR was performed on each colony every 24 hours for 14 days to amplify the TNR tract (see Fig. 6 for cartoon). **Left panel:** The tracts from time course E4 were resolved on a 12% polyacrylamide gel and stained with EtBr. The numbers across the top of the gel indicate the day of the time course. **Right panel:** Southern blot of the gel in **left panel** to demonstrate that the expansion products contain TNR sequence. The lanes marked **C** in each panel indicate the 75bp tract amplified from the TNR plasmid control. B. **msh3Δ:** Expansion events were selected in the msh3Δ \((CTG)_{25}\) background and confirmed by PCR. Individual colonies were followed over several days; colony PCR was performed on each colony every 24 hours for 10 days to amplify the TNR tract. **Top panel:** The tracts from time course V1 were resolved on a 12% polyacrylamide gel and strained with Et Br. The numbers across the top of the gel indicate the day of the time course. **Lower left panel:** The tracts from time course C8 were resolved on a 12% polyacrylamide gel and stained with EtBr. The numbers across the top of the gel indicate the day of the time course. **Lower right panel:** Southern blot of the gel in **lower left panel** to demonstrate that the expansion products indicated by the arrows contain TNR sequence. The lanes marked **C** in each panel indicate the 75bp tract amplified from the TNR plasmid control.

**Figure 8. Colony time courses in MSH3 and msh3Δ backgrounds.** The TNR tract size was plotted as a function of time to demonstrate trends in the data. Each panel represents an independent \((CTG)_{25}\) integrant in the different strain backgrounds. Each curve in a single panel represents the time course of an independent expansion event in that background. A.
**MSH3**: All *MSH3* colony time courses were plotted. Each plot show the tract size changes in independent colonies from a single *MSH3(CTG)_{25}* isolate. A total of 64 time courses were performed: 28 showed progressive expansion, 3 showed tract stability, 11 showed no amplification (loss of tract) and 22 showed high background with multiple PCR products. This last category was difficult to interpret but exhibited trends similar to those analyzed in Fig. 8A; 9 tracts exhibited increasing trends and 4 tracts appeared stable. **B msh3Δ**: All *msh3Δ* time courses were plotted and linear regressions were calculated. A total of 64 time courses were performed: 0 showed progressive expansion, 23 showed tract stability, 7 tracts showed contractions, 26 showed no amplification (loss of tract) and 6 showed high background with multiple PCR products. Linear regressions were plotted for each time course to illustrate the general trends in tract length.

**Figure 9. Sizes of each step (tract size change) in colony time course experiments.** Each progressive tract length change, defined as the loss of one tract length and the concomitant appearance of a different tract length, was plotted as a function of size for *MSH3* (A) and *msh3Δ* (B) using GraphPad Prism. The median and interquartile range for each step is indicated in black. **A. MSH3**: The initial expansion size selected on 5-FOA is indicated (pink circles). The orange squares indicate the size of each individual tract upon the first tract length change, or step. Similarly, the second, third and fourth steps are shown (olive triangles, green inverted triangles, blue diamonds, respectively). The median size of each step is essentially unchanged. **B. msh3Δ**: The sizes of the initial expansion events are shown (pick circles); the median is comparable to that of the *MSH3* time courses. Only one progressive step was observed (orange squares) and this was a step down to a short tract length.

**Figure 10. Schematic of the microcolony time course experiment protocol.** Individual colonies with TNR tract expansions were selected on plates containing 5-FOA; the tract increase was confirmed by colony PCR in each case. Instead of initiating the time course with
a potentially mixed population, the microcolony time course was started with single cells from a colony with an expanded tract. These cells were restricted to between 8 and 10 cell divisions, with an eye to minimizing the heterogeneity within the population. See Materials and Methods for additional details.

**Figure 11. Dynamic changes in TNR tracts starting from a single cell.** Expansion events were selected in either the *MSH3 (CTG)*₂₅ (A) or *msh3Δ (CTG)*₂₅ (B) background and confirmed by PCR. Individual cells from these colonies were isolated and allowed to undergo 8-10 rounds of replication, resulting in a microcolony approximately 250-1000 cells in size. A single cell was then taken from this microcolony to propagate another microcolony. The remainder was used to amplify the TNR tract by PCR to determine tract length (see Fig. 10 for cartoon).  

A. An example of TNR tract dynamics in a *MSH3* microcolony (F from Table 2). There are multiple changes in the TNR tract over time.  

B. Two examples of TNR tract dynamics in an *msh3Δ* microcolony (N and Q, respectively, from Table 3). The red asterisk indicates the position of the initial expansion tract size, as determined by Southern blot. The upper bands in the lower panel do not contain TNR tract sequences, as determined by Southern blot (data not shown). The tract appears more stable than in the *MSH3* background. Both gels are 12% polyacrylamide gels stained with EtBr; the images of been inverted for ease of viewing. The lanes marked C in each panel indicate the 75bp tract amplified from the TNR plasmid control. The numbers across the top of the gels indicate the time point.
Table 1. Summary of rates of tract length changes calculated from liquid, colony and microcolony time courses

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Rates of tract length changes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Liquid a (95% C.I.) b</td>
</tr>
<tr>
<td>MSH3(CTG)</td>
<td>1.2 x 10⁻³ (9.0 x 10⁻⁴ – 1.6 x 10⁻³)</td>
</tr>
<tr>
<td>MSH3(CAG)</td>
<td>3.1 x 10⁻³ (2.4 x 10⁻³ – 3.4 x 10⁻³)</td>
</tr>
<tr>
<td>msh3Δ(CTG)</td>
<td>3.1 x 10⁻⁴ (1.3 x 10⁻⁴ – 6.0 x 10⁻⁵)</td>
</tr>
<tr>
<td>msh3Δ(CAG)</td>
<td>1.3 x 10⁻⁴ (2.9 x 10⁻⁵ – 3.7 x 10⁻⁵)</td>
</tr>
<tr>
<td>MSH3(CTG)₂₅</td>
<td>&lt;5 x 10⁻⁵ h</td>
</tr>
<tr>
<td>MSH3(CAG)₂₅</td>
<td>5.1 x 10⁻⁵ h</td>
</tr>
</tbody>
</table>

a expansions only  

b 95% confidence intervals determined by F statistic  

c not determined  

d predominantly contractions  

e slight bias toward expansions  

f bias toward contractions  

g no tract changes observed  

h low numbers precluded accurate calculation of 95% C.I.
Table 2. *MSH3* Microcolony Tract Dynamics

<table>
<thead>
<tr>
<th>Day</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F^a</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>M</th>
<th>R</th>
<th>P</th>
<th>S</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial tract</td>
<td>32</td>
<td>30</td>
<td>32</td>
<td>33</td>
<td>31</td>
<td>30</td>
<td>31</td>
<td>32</td>
<td>34</td>
<td>30</td>
<td>30</td>
<td>31</td>
<td>33</td>
<td>32</td>
<td>31</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>-5</td>
<td>0</td>
<td>-2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>-3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-7</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>-2</td>
<td>0</td>
<td>-2</td>
<td>-3</td>
<td>-2</td>
<td>-2</td>
<td>0</td>
<td>-3</td>
<td>5</td>
<td>0</td>
<td>-4</td>
<td>-3</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>-3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0</td>
<td>-3</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-4</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>-2</td>
<td>0</td>
<td>-7</td>
<td>0</td>
<td>8</td>
<td>-5</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>-1</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>7</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-7</td>
<td>-2</td>
<td>6</td>
<td>-8</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>-4</td>
<td>-2</td>
<td>0</td>
<td>-2</td>
<td>0</td>
<td>0</td>
<td>-2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>-</td>
<td>-1</td>
<td>0</td>
<td>-4</td>
<td>0</td>
<td>4</td>
<td>14</td>
<td>3</td>
<td>5</td>
<td>0</td>
<td>4</td>
<td>-3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Final tract</td>
<td>46</td>
<td>27</td>
<td>32</td>
<td>30</td>
<td>33</td>
<td>40</td>
<td>38</td>
<td>46</td>
<td>37</td>
<td>42</td>
<td>30</td>
<td>35</td>
<td>30</td>
<td>32</td>
<td>31</td>
</tr>
<tr>
<td>Net change</td>
<td>14</td>
<td>-3</td>
<td>0</td>
<td>-3</td>
<td>2</td>
<td>10</td>
<td>7</td>
<td>14</td>
<td>3</td>
<td>12</td>
<td>0</td>
<td>4</td>
<td>-3</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

^a corresponds to gel in Figure 11A.

^b Initial expansion size following selection in the presence of 5-FOA, in number of repeats.

^c Change in number of repeats in tract.

^d No amplification of tract.

^e Number of repeats.

^f Net change in number of repeats within tract.
Table 3. \textit{msh3Δ} Microcolony Tract Dynamics

| Day | A   | B   | C   | D   | E   | F   | G   | H   | I   | J   | K   | L   | M   | N*  | O   | P   | Q2  | R   | S   | T   | U   | V   | W   |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Initial tract | 30  | 31  | 30  | 31  | 32  | 32  | 32  | 29  | 30  | 30  | 32  | 30  | 30  | 36  | 29  | 29  | 30  | 30  | 30  | 29  | 29  | 30  | 30  |
| 1   | 0   | 0   | 0   | 0   | -3  | -2  | -2  | 0   | 0   | 0   | -2  | -1  | 0   | -    | 0   | 0   | 0   | 0   | 0   | -2  | 0   |    |
| 2   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | -7  | 0   | 0   | 0   | 0   | 0   | 0   | 0   |    |
| 3   | -   | -   | -   | 0   | -5  | -5  | -5  | 0   | 0   | 0   | 0   | 0   | 0   | -2  | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   |    |
| 4   | -   | -   | -   | 0   | 0   | 0   | -1  | -   | 0   | 0   | 0   | 5   | -   | 0   | -   | 0   | -   | 0   | -   | 0   | -   |    |
| 5   | -   | -   | -   | -   | -   | -   | -   | -   | -   | -   | 0   | -   | 0   | -   | 0   | -   | 0   | 0   | -   |    |    |    |
| 6   | 0   | 0   | 0   | 0   | -5  | 0   | 5   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | -   | 0   | -   | 0   |    |    |
| 7   | 0   | 0   | 0   | 0   | -1  | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | -   | 0   | 0   | 0   | 0   | 0   | -   |    |
| 8   | -1  | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | -   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 2   | 6   |    |
| 9   | 0   | 0   | 0   | 0   | -4  | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | -   | 1   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   |
| 10  | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | -   | 1   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | -4  |
| 11  | 0   | 0   | 0   | 0   | 1   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | -   | 0   | -   | 0   | 0   | 0   | 0   | 0   | 0   | -4  |
| Final tract | 29  | 31  | 30  | 30  | 31  | 29  | 30  | 30  | 30  | 29  | 29  | 30  | 29  | 30  | 29  | 29  | 30  | 29  | 30  | 29  | 31  | 28  |    |
| Net change | -1  | 0   | 0   | 0   | 0   | -3  | -2  | -1  | 0   | 0   | 0   | 0   | -2  | -1  | 0   | -7  | 0   | 0   | 0   | 0   | 0   | 0   | 0   | -2  |

\(a\) corresponds to gel in upper panel of Figure 11B.

\(b\) corresponds to gel in lower panel of Figure 11B.

\(c\) Initial expansion size following selection in the presence of 5-FOA, in number of repeats.

\(d\) Change in number of repeats in tract.

\(e\) No amplification of tract.

\(f\) Number of repeats.

\(g\) Net change in number of repeats within tract.
Table 4. Trinucleotide repeat tract dynamics in microcolony time courses

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Type of Event</th>
<th># Events</th>
<th>Median size of event (range in # of repeats)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSH3 (CTG)(^a)</td>
<td>Expansion</td>
<td>38</td>
<td>3 (1-14)</td>
</tr>
<tr>
<td></td>
<td>Contraction</td>
<td>30</td>
<td>3 (1-8)</td>
</tr>
<tr>
<td></td>
<td>Final TNR Tract Length</td>
<td>33 (27-46)</td>
<td></td>
</tr>
<tr>
<td>msh3(\Delta) (CTG)(^b)</td>
<td>Expansion</td>
<td>10</td>
<td>3 (1-6)</td>
</tr>
<tr>
<td></td>
<td>Contraction</td>
<td>18</td>
<td>2 (1-7)</td>
</tr>
<tr>
<td></td>
<td>Final TNR Tract Length</td>
<td>30 (29-32)</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) 96 total time courses were initiated: 32 time courses were completed to at least day 12; of these, 13 exhibited TNR tract dynamics, 2 exhibited a stable tract length, 15 exhibited no tract amplification and 2 were uninterpretable due to high background.

\(^b\) 105 total time courses were initiated: 37 time course were completed to at least day 10); of these 14 exhibited TNR tract dynamics, 9 exhibited a stable tract length, 14 exhibited no tract amplification.
Williams and Surtees Figure 1
Williams and Surtees Figure 2
A. *MSH3* logarithmic

B. *msh3A* logarithmic

<table>
<thead>
<tr>
<th>Genotype</th>
<th>% Expanded (Day 7)</th>
<th>% Expanded (Day 14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>*MSH3 (CT9)*₁₀</td>
<td>40% (8/20)</td>
<td>45% (9/20)</td>
</tr>
<tr>
<td>*MSH3 (CT9)*₅</td>
<td>45% (9/20)</td>
<td>45% (9/20)</td>
</tr>
<tr>
<td>*MSH3 (CT9)*₁</td>
<td>45% (9/20)</td>
<td>50% (10/20)</td>
</tr>
<tr>
<td>Total</td>
<td>43% (28/66)</td>
<td>47% (23/48)</td>
</tr>
<tr>
<td>*MSH3 (CA9)*₁₀</td>
<td>37% (11/30)</td>
<td>70% (14/20)</td>
</tr>
<tr>
<td>*MSH3 (CA9)*₅</td>
<td>47% (14/30)</td>
<td>85% (17/20)</td>
</tr>
<tr>
<td>Total</td>
<td>42% (26/62)</td>
<td>78% (31/40)</td>
</tr>
</tbody>
</table>

Williams and Surtees

Figure 4
### A. MSH3 stationary

<table>
<thead>
<tr>
<th>Genotype</th>
<th>% Expanded (Day 7)</th>
<th>% Expanded (Day 14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSH3 (CT9)_a</td>
<td>0% (0/20)</td>
<td>0% (0/20)</td>
</tr>
<tr>
<td>MSH3 (CT9)_b</td>
<td>10% (2/20)</td>
<td>10% (2/20)</td>
</tr>
<tr>
<td>MSH3 (CT9)_c</td>
<td>10% (2/20)</td>
<td>15% (3/20)</td>
</tr>
<tr>
<td>Total</td>
<td>7% (4/80)</td>
<td>8% (6/80)</td>
</tr>
<tr>
<td>MSH3 (CA9)_a</td>
<td>5% (1/20)</td>
<td>5% (1/20)</td>
</tr>
<tr>
<td>MSH3 (CA9)_b</td>
<td>15% (3/20)</td>
<td>25% (5/20)</td>
</tr>
<tr>
<td>MSH3 (CA9)_c</td>
<td>15% (3/20)</td>
<td>15% (3/20)</td>
</tr>
<tr>
<td>Total</td>
<td>12% (7/80)</td>
<td>16% (8/80)</td>
</tr>
</tbody>
</table>

### B. msh3Δ stationary

<table>
<thead>
<tr>
<th>Genotype</th>
<th>% Expanded (Day 7)</th>
<th>% Expanded (Day 14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>msh3Δ (CT9)_a</td>
<td>10% (2/20)</td>
<td>0% (0/20)</td>
</tr>
<tr>
<td>msh3Δ (CT9)_b</td>
<td>5% (1/20)</td>
<td>5% (1/20)</td>
</tr>
<tr>
<td>msh3Δ (CT9)_c</td>
<td>10% (2/20)</td>
<td>0% (0/20)</td>
</tr>
<tr>
<td>Total</td>
<td>8% (6/80)</td>
<td>1% (1/80)</td>
</tr>
<tr>
<td>msh3Δ (CA9)_a</td>
<td>0% (0/20)</td>
<td>0% (0/20)</td>
</tr>
<tr>
<td>msh3Δ (CA9)_b</td>
<td>0% (0/20)</td>
<td>0% (0/20)</td>
</tr>
<tr>
<td>msh3Δ (CA9)_c</td>
<td>0% (0/20)</td>
<td>0% (0/20)</td>
</tr>
<tr>
<td>Total</td>
<td>0% (0/80)</td>
<td>0% (0/80)</td>
</tr>
</tbody>
</table>

*Williams and Surtees Figure 5*
Colony PCR to confirm expansion

5-FOA

Transfer colonies with expanded tract

SC-nls

24 hours (~10 generations)

SC-nls

Colony PCR to assess tract size

24 hours (~10 generations)

SC-nls

Colony PCR to assess tract size

Williams and Surtees
Figure 6

Continued through Day 14
A. *MSH3* colony time course

B. *msh3Δ* colony time courses

Williams and Surtees Figure 7
A. *MSH3* colony time courses

B. *msh3Δ* colony time course

Williams and Surtees
Figure 8
A. *MSH3(CTG) colony steps*

B. *ms h3 Δ(CTG) colony steps*

Williams and Surtees Figure 9
A. *MSH3* microcolony time course

B. *msh3Δ* microcolony time courses

Williams and Surtees Figure 11